

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 15:09:06 ; Search time 1481.25 Seconds
(without alignments)
8990.874 Million cell updates/sec

Title: US-09-308-080-1

Perfect score: 861
Sequence: 1 TGTATATGAGATAATATAT.....AGTGGCAATATATTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_rod: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_vi: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_v11: *
59: gb_v12: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_r01: *
95: gb_r02: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	773.6	89.8	180286	80	AL356574 Homo sapi
2	607	70.5	626	97	HSU57655 Human dihyd
3	486.6	56.5	854	93	HSXP08GEN X95670 H.sapiens D
4	172.8	20.1	3951	97	HSU09178 U09178 Human dihyd
5	172.8	20.1	3957	9	AR027740 Sequence
6	171.2	19.9	3078	85	AB003063 Homo sapi
7	171.2	19.9	4409	97	HSU20938 U20938 Human lymph
8	154.8	18.0	4421	7	BTU20981 Bos taurus

9	151.6	17.6	4441	7	SSU09179	U09179	Sus scrofa
10	151.6	17.6	4447	9	AR027741	AR027741	Sequence
11	146.8	17.0	4358	94	D85035	D85035	Rattus norv
12	72	8.3	43803	6	CEL25F6	U39742	Caenorhabdi
13	71.4	8.3	24429	63	AC014157	AC014157	Drosophila
14	71.4	8.3	303092	4	AE003446	AE003446	Drosophila
15	69.8	8.1	2346	6	DM05491	U65491	Drosophila
16	56.8	6.6	117674	84	DMBR30C13	AL122025	Drosophila
17	51	5.9	141953	71	AC044793	AC044793	Homo sapi
18	50	5.8	156060	60	AC004153	AC004153	Plasmodiu
19	49.8	5.8	150803	61	AC010228	AC010228	Homo sapi
20	49.6	5.8	1017	6	DDCPR02U	X04775	Dicystosteli
21	49.4	5.7	147490	90	AL445207	AL445207	Human DNA
22	49.4	5.7	160728	70	AC026816	AC026816	Homo sapi
23	49.4	5.7	161025	66	AC021276	AC021276	Homo sapi
24	49.4	5.7	164450	82	AP000794	AP000794	Homo sapi
25	49.2	5.7	168674	69	AC025645	AC025645	Homo sapi
26	48.8	5.7	194892	15	T14P8	AF069298	Arabidops
27	48.8	5.7	194892	13	ATCHR1V6	AL162494	Arabidops
28	48.8	5.7	217327	79	AL162723	AL162723	Homo sapi
29	48.2	5.6	1149	5	AF044859	AF044859	Eurema da
30	48.2	5.6	12029	4	AE001412	AE001412	Plasmodiu
31	48	5.6	590	96	PARMTD14D	K00907	paramecium
32	48	5.6	692	96	PARMTD14C	K00906	paramecium
33	48	5.6	726	96	PARMTD14C	K00906	paramecium
34	48	5.6	760	96	PARMTD14A	K00904	paramecium
35	48	5.6	40469	6	MIPAGEN	X15917	Paramecium
36	48	5.6	47573	5	AF030694	AF030694	Plasmodiu
37	47.4	5.5	112519	85	AB041992	AB041992	Homo sapi
38	47.4	5.5	148965	72	AC060784	AC060784	Homo sapi
39	47.4	5.5	165823	71	AC027625	AC027625	Homo sapi
40	47.4	5.5	177920	64	AC016700	AC016700	Homo sapi
41	47.4	5.5	300000	91	AP002529	AP002529	Homo sapi
42	47.2	5.5	86827	96	PFMAL13P5	AL034556	Plasmodiu
43	47	5.5	7218	10	166494	166494	Sequence 14
44	47	5.5	65691	96	PFMAL13P1	297348	Plasmodiu
45	46.8	5.4	87509	12	AC004482	AC004482	Arabidops

ALIGNMENTS

RESULT 1
LOCUS AL356574/c
DEFINITION Homo sapiens chromosome 1 clone RP11-359C24, *** SEQUENCING IN
PROGRESS ***
ACCESSION AL356574
VERSION AL356574.3 GI:9797568
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 180286)
PlumB.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9213933.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA359C24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 166105 bases at least Q40

Consensus quality: 171672 bases at least Q30
Consensus quality: 174563 bases at least Q20
Insert size: 177786; sum-of-contigs
Insert size: 186027; 5.1% error; agarose-gel
Quality coverage: 3.18x in Q20 bases; sum-of-contigs quality
coverage: 3.10x in Q20 bases; agarose-gel

NOTE: This is a 'working draft' sequence. It currently
consists of 26 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 15090: contig of 15090 bp in length
15091 15190: gap of 100 bp
15191 23545: contig of 8355 bp in length
23546 23645: gap of 100 bp
23646 26766: contig of 3121 bp in length
26767 26866: gap of 100 bp
26867 38453: contig of 11587 bp in length
38454 38553: gap of 100 bp
38554 45878: contig of 7325 bp in length
45879 45978: gap of 100 bp
45979 51984: contig of 6006 bp in length
51985 52084: gap of 100 bp
52085 58204: contig of 6120 bp in length
58205 58304: gap of 100 bp
58305 71630: contig of 13376 bp in length
71631 71730: gap of 100 bp
71731 75338: contig of 3608 bp in length
75339 75438: gap of 100 bp
75439 85213: contig of 9775 bp in length
85214 85313: gap of 100 bp
85314 88517: contig of 3204 bp in length
88518 88617: gap of 100 bp
88618 92454: contig of 3837 bp in length
92455 92554: gap of 100 bp
92555 98200: contig of 5646 bp in length
98201 98300: gap of 100 bp
98301 111029: contig of 12729 bp in length
111030 111129: gap of 100 bp
111130 117311: contig of 6182 bp in length
117312 117411: gap of 100 bp
117412 121093: contig of 3682 bp in length
121094 121193: gap of 100 bp
121194 137336: contig of 16143 bp in length
137337 137436: gap of 100 bp
137437 141486: contig of 4050 bp in length
141487 141586: gap of 100 bp
141587 146703: contig of 5117 bp in length
146704 146803: gap of 100 bp
146804 149034: contig of 2231 bp in length
149035 149134: gap of 100 bp
149135 154764: contig of 5630 bp in length
154765 154864: gap of 100 bp
154865 159990: contig of 5126 bp in length
159991 160090: gap of 100 bp
160091 163742: contig of 3652 bp in length
163743 163842: gap of 100 bp
163843 166322: contig of 2480 bp in length
166323 166422: gap of 100 bp
166423 176700: contig of 10278 bp in length
176701 176800: gap of 100 bp
176801 180286: contig of 3486 bp in length.

FEATURES

Source

1. 180286
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-359C24"
/clone_id="RP11-1.2"

misc_feature	1. .15090	/note="assembly_fragment:00687
	clone_end:sp6	
	vector_side:left"	
misc_feature	15191. .23545	/note="assembly_fragment:00038
	fragment_chain:1"	
misc_feature	23646. .26766	/note="assembly_fragment:00972
	fragment_chain:1"	
misc_feature	26857. .38453	fragment_chain:1"
	/note="assembly_fragment:01030	
misc_feature	38554. .45878	fragment_chain:1"
	/note="assembly_fragment:00178	
misc_feature	45979. .51984	fragment_chain:2"
	/note="assembly_fragment:00498	
misc_feature	52085. .58204	fragment_chain:2"
	/note="assembly_fragment:00217	
misc_feature	58305. .71630	fragment_chain:3"
	/note="assembly_fragment:01029	
misc_feature	71731. .75338	fragment_chain:3"
	/note="assembly_fragment:00265	
misc_feature	75349. .85213	fragment_chain:4"
	/note="assembly_fragment:01189	
misc_feature	85314. .88517	fragment_chain:4"
	/note="assembly_fragment:00320	
misc_feature	88618. .92454	fragment_chain:5"
	/note="assembly_fragment:00693	
misc_feature	92555. .98200	fragment_chain:5"
	/note="assembly_fragment:00834	
misc_feature	98301. .111029	fragment_chain:6"
	/note="assembly_fragment:00385	
misc_feature	111130. .117311	fragment_chain:6"
	/note="assembly_fragment:00932	
misc_feature	117412. .121093	fragment_chain:7"
	/note="assembly_fragment:01129	
misc_feature	121194. .137336	fragment_chain:7"
	/note="assembly_fragment:00160"	
misc_feature	137437. .141486	/note="assembly_fragment:00194"
	/note="assembly_fragment:00234"	
misc_feature	141587. .146703	/note="assembly_fragment:00234"
	/note="assembly_fragment:00529"	
misc_feature	149135. .154764	/note="assembly_fragment:00590"
	/note="assembly_fragment:00590"	
misc_feature	154865. .159900	/note="assembly_fragment:01001"
	/note="assembly_fragment:01119"	
misc_feature	160091. .163742	/note="assembly_fragment:01294"
	/note="assembly_fragment:01294"	
misc_feature	166423. .176700	/note="assembly_fragment:01323"
	/note="assembly_fragment:01326"	
misc_feature	2522 others	

Query Match	89.8%	Score 773.6	DB 80	Length 180286
Best Local Similarity	95.7%	Pred. No. 2.2e-141		
Matches 827	Conservative 0	Mismatches 34	Indels 3	Gaps 3
Qy 1	TGTATATGAGATTAATATTTTGTGTTTTTCCCTGTCTTAACCTTAGGTTCAAGAAAGT	60		
Db 50426	TGTATATGAGATTAATATTTTATTTTATTTTGGCTATTTCAAACTTAGAATTTCAAGAACT	50367		
Qy 61	AATTTATCTGGAGCTAACAAATACCTTTATTTTACCTTTTATTTTGAAGTAGTTATGTT	120		
Db 50366	AATTTAGCTAAAGCTAACAAATACCTTTATTTTACCTTTTATTTTGAAGTAGTTATGTT	50307		
Qy 121	CAATTCATATTTAATGTATATTTAAAAATTCCTGTGCAATATGTAGAGGAGCACTCATAT	180		
Db 50306	CAATTCATATTTAATGTATATTTAAAAATTCCTGTGCAAAATGTAGAGGAGCACTCATAT	50247		
Qy 181	AAATTTGTGATATGGAATATGAGCATTAATTAAGATTAATACCTTTCTGTGCAAAAGG	240		
Db 50246	AAATTTGTGATATGGAATATGAGCATTAATTAAGATTAATACCTTTCTGTGCAAAAGG	50187		
Qy 241	AGACTCAATATCTTACTCTTTTCATGAGAGACATTTGTGACAAATGTGTTCCCTCATATCAT	300		
Db 50186	AGACTCAATATCTTACTCTTTTCATGAGAGACATTTGTGACAAATGTGTTCCCTCATATCAT	50127		
Qy 301	CCGGGGAACCACTCTGTGGCCCATGTATATGGCCCTGTGACAAAGCTCCTTTTCTGAATATGGA	360		
Db 50126	CCGGGGAACCACTCTGTGGCCCATGTATATGGCCCTGTGACAAAGCTCCTTTCTGAATATGGA	50067		
Qy 361	GCTCATCATGTGAGAAAAGGGCTGCTATTTGGTGTCGAAAGTGCACCTGAACCTTAAGAGCTGA	420		
Db 50066	GCTCATCATGTGAGAAAAGGGCTGCTATTTGGTGTCGAAAGTGCACCTGAACCTTAAGAGCTGA	50007		
Qy 421	CTTCCAGACACAATTAAGTGTGA-TAAAAATCTAAACAAGAAATTTGGCATTAAGTTGGT	479		
Db 50006	CTTTCAGACACAATTAAGTGTGAATTAACATCTTAACAACAAGAAATTTGGCATTAAGTTGGT	49947		
Qy 480	GAATGTTTATTTAAACATCCCAATTCATATAGGCTTATTAATATTTAATGTATATTTATCA	539		
Db 49946	GAATGTTTATTTAAACATCCCAATTCATATAGGCTTATTAATATTTAATGTATATTTATATA	49887		
Qy 540	ACGAATCTGCCAGTTGCTTTGCTGATGATAGAAAATTAAGATTAAGAAAAGAAAAGCTCAAG	599		
Db 49886	AAGAATCTGCCAGTTGCTTTGCTGATGATAGAAAATTAAGATTAAGAAAAGAAAAGCTCAAG	49827		
Qy 600	AACTCATAAAAAACCACACAATGTGAAGCTCTGTTATTAATGGGTGCCATGTAAAGATGGA	659		
Db 49826	AACTCATAAAAAACCACACAATGTGAAGCTTTGTTATTAATGGGTGCCATGTAAAGATGGA	49767		
Qy 660	AGAAGTATCTCATATAGCAGAGAAGAGAAATGAATATCTCATTTTATGTAGTTGGCCC	719		
Db 49766	AGAAGTATCTCATATAGCAGAGAAGAGAAATGAATATCTCATTTTATGTAGTTGGGTTT	49707		
Qy 720	CCACGTATATGGCGGTGATTTATATAGGTGATGAGC-CAGGAACAATTTGTAACCTATA	778		
Db 49706	TCACGTATATGGCGGTGATTTATATAGGTGATGAGCCTTAGAGAAATTTGTAACCTATA	49647		
Qy 779	AACCACTCAAAATATTAACCCGAGGACAGAGCAGCATATC-TCCTATGAAAGCCTGTATTT	837		
Db 49646	AATCATTTTAAATATTAACCTGAGGACAGAGCAGCATATCTTCTCATGAAAGCTATATTT	49587		
Qy 838	ACTCAGTGGGAATATTTATTTAA 861			
Db 49586	TTTTCAGTGGGAATATTTATTTAA 49563			

[illegible]

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS Fernandez-Salguero, P.M., Wei, X., Jones, S., Idle, J.R. and Gonzalez, F.J.
TITLE Lack of dihydropyrimidine dehydrogenase activity and thymine-uraciluria caused by a common splice mutation Hum. Mutat. (1996) In press
JOURNAL 2 (bases 1 to 626)
AUTHORS Fernandez-Salguero, P.
REFERENCE Direct Submission
Submitted (07-MAY-1996) Pedro M. Fernandez-Salguero, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA
TITLE
JOURNAL
FEATURES
source
1. 626
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
179..343
/gene="DYPD"
CDS
<179..>343
/gene="DYPD"
/function="uracil and thymine metabolism"
/codon_start=1
/evidence=experimental
/product="dihydropyrimidine dehydrogenase"
/protein_id="AA07049.1"
/db_xref="GI:1518607"
/translation="DIVTNSPRIIRGTTSGPMYGRGSSFLNIELISEKTAAYWCOS VTELKADFPDN"
BASE COUNT 221 a 100 c 117 g 188 t
ORIGIN

Query Match 70.5% Score 607 DB 97 Length 626
Best Local Similarity 99.0% Pred. No. 6.8e-109
Matches 621: Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 90 TTTCCTTTTATTTGCAAGTAGTTATGTCATTCATTTAATTAATTAATAAATT 149
|||||
Db 1 TTTCCTTTTATTTGCAAGTAGTTATGTCATTCATTTAATTAATTAATAAATT 60
|||||

QY 150 CCTCTGCAAAATATGTGAGGAGGACCTCATAAATATTTGTCATATGAAATGAGCAGATA 209
|||||
Db 61 CCTCTGCAAAATATGTGAGGAGGACCTCATAAATATTTGTCATATGAAATGAGCAGATA 119
|||||

QY 210 ATTAAGATTATAGCTTTCTTGTGCAAAAGAGACTCAATATCTTACTCTTCATGAG 269
|||||
Db 120 ATTAAGATTATAGCTTTCTTGTGCAAAAGAGACTCAATATCTTACTCTTCATGAG 179
|||||

QY 270 ACATTGTGACAAATGTTCCCCATTAATCATCCGGGGAACACCTGCGCCCATGTATG 329
|||||
Db 180 ACATTGTGACAAATGTTCCCCATTAATCATCCGGGGAACACCTGCGCCCATGTATG 239
|||||

QY 330 GCCCTGGAACAAAGCTCTTTCTGAAATATTTGAGCTCATGAGAGAAAAGGCTGCATATT 389
|||||
Db 240 GCCCTGGAACAAAGCTCTTTCTGAAATATTTGAGCTCATGAGAGAAAAGGCTGCATATT 299
|||||

QY 390 GGTTCTCAAAAGTGTACGTAAGCTGAAGCTTCCCGACAGACGTAAGTGTGATAAAAA 449
|||||
Db 300 GGTTCTCAAAAGTGTACGTAAGCTGAAGCTTCCCGACAGACGTAAGTGTGATAAAAA 359
|||||

QY 450 TCTTAAACAAAGAGATTGGCATTAAGTTGGTAATGTTAATTAACATCCATTCATAGG 509
|||||
Db 360 TCTTAAACAAAGAGATTGGCATTAAGTTGGTAATGTTAATTAACATCCATTCATAGG 419
|||||

QY 510 CTTATTAATATTAATGTTGATATTTTATCAAGAAATCTGCCAGTTGCTTGGCTGATGAT 569
|||||
Db 420 CTTATTAATATTAATGTTGATATTTTATCAAGAAATCTGCCAGTTGCTTGGCTGATGAT 479
|||||

QY 570 AGAAAGTAAAAAAGAAAGAAAGCTCAAGAACTCATATAAAACCCACACATGTGAAGCT 629
|||||
Db 480 AGAAAGTAAAAAAGAAAGAAAGCTCAAGAACTCATATAAAACCCACACATGTGAAGCT 539
|||||

QY 630 CTTTATTAATGCGTGCATGTAAGATGAGAGATATCTACATTAAGCAGAGAGAGA 689
|||||
Db 540 CTTTATTAATGCGTGCATGTAAGATGAGAGATATCTACATTAAGCAGAGAGAGA 599
|||||

QY 690 AATGAATATCTCATTTTATTTGAGTTGG 716
|||||
Db 600 AATGAATATCTCATTTTATTTGAGTTGG 626
|||||

RESULT 3
HSDYDPGEN 854 bp DNA PRI 17-FEB-1997
LOCUS H.sapiens DYPD gene, partial sequence.
DEFINITION X95670
ACCESSION X95670.1 GI:1246752
VERSION dihydropyrimidine dehydrogenase; DYPD gene; exon X; mutation.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 854)
AUTHORS Vreken, P., Van Kullenburg, A.B., Meinsma, R., Smit, G.P., Bakker, H.D., De Abreu, R.A. and van Genailp, A.H.
TITLE A point mutation in an invariant splice donor site leads to exon skipping in two unrelated Dutch patients with dihydropyrimidine dehydrogenase deficiency
J. Inher. Metab. Dis. 19 (5), 645-654 (1996)
JOURNAL 97047101
MEDLINE 2 (bases 1 to 854)
REFERENCE Vreken, P.
AUTHORS Direct Submission
TITLE Submitted (14-FEB-1996) P. Vreken, Academic Med.Center, Univ. of Amsterdam, Dep. of Pediatrics and Clinical Chemistry, PO-224, P.O. Box 22700, NL-1100 DE Amsterdam, NETHERLANDS
COMMENT Related sequences:-
Meinsma et al, DNA Cell Biol. 14:1-6 (1995), U20938, and Yokota H. et al, J. Biol. Chem. 269:23192-23196 (1994).
FEATURES
source
1. 854
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="liver"
/map="p22"
<1..568
intron
569..734
gene
/gene="DYPD"
CDS
<569..>734
/gene="DYPD"
/EC_number="1.3.1.2"
/note="exon X"
/codon_start=2
/product="dihydropyrimidine dehydrogenase (NADP+)"
/protein_id="CA64973.1"
/db_xref="GI:1246753"
/db_xref="SWISS-PROT:Q12882"
/translation="DIVTNSPRIIRGTTSGPMYGRGSSFLNIELISEKTAAYWCOS VTELKADFPDN"
exon
569..734
725
/gene="DYPD"
variation
/gene="DYPD"
/replace="C"
735..>854
intron
735
variation
/note="changes the splice donor sequence and leads to exon skipping"
/replace="A"
BASE COUNT 301 a 126 c 130 g 297 t
ORIGIN

Query Match	56.5%;	Score 486.6;	DB 93;	Length 854;
Best Local Similarity	96.0%;	Pred. No. 2e-85;		
Matches 531;	Conservative 0;	Mismatches 19;	Indels 3;	Gaps 3;
QY	1	TGTTAATGAAGATTAATATTTTGTGTTTGGCTGTTCTAAACCTAGGCTTACAAAGT	60	
DB	304	TGTTAATGAATTAATATTTTGTGCTATTCCTAAACCTAGATTAACAAAGT	363	
QY	61	AATTTATCTGAGCTACCAATCTTATTTTACCTTTTATTTGCAAGTAGTTATGTT	120	
DB	364	AATTTATCTGAGCTACCAATCTTATTTTACCTTTTATTTGCAAGTAGTTATGTT	422	
QY	121	CAATTCAATTTATGATATTAATAATTCCTGCAAAATATGAGAGGAGCCTGATA	180	
DB	423	CAATTCAATTTATGATATTAATAATTCCTGCAAAATATGAGAGGAGCCTGATA	482	
QY	181	AAATATTTGATATGAATGAGCAGATTAATAAGATTATAGCTTTCTTTGCAAAAG	240	
DB	483	AAATA-TGTCATATGAAATGAGCAGATTAATAAGATTATAGCTTTCTTTGCAAAAG	541	
QY	241	AGACTCAATCTTTACTCTTTCATGAGACATTTGACAAATGTTCCCCCATATCAT	300	
DB	542	AGACTCAATCTTTACTCTTTCATGAGACATTTGACAAATGTTCCCCCATATCAT	601	
QY	301	CCGGGGAACACCTCTGGCCCATGATGGCCCTGACAAAGCTCCTTCTGAAATATTGA	360	
DB	602	CCGGGGAACACCTCTGGCCCATGATGGCCCTGACAAAGCTCCTTCTGAAATATTGA	661	
QY	361	GCTCATAGTGAAGAAAGCGCTGCATATTTGTCGCAAGTGCACGACTAAAGGCTGA	420	
DB	662	GCTCATAGTGAAGAAAGCGCTGCATATTTGTCGCAAGTGCACGACTAAAGGCTGA	721	
QY	421	CTTCCGACGACATAGTGTGA-TAAATCTTAAACAGACAAATGGCATAAGTGT	479	
DB	722	CTTCCGACGACATAGTGTGA-TAAATCTTAAACAGACAAATGGCATAAGTGT	781	
QY	480	GAATGTTTATTAACATCCATTCATAGGCTTATAATATTAATGATGATTTATTTCA	539	
DB	782	GAATGTTTATTAACATCCATTCATAGGCTTATAATATTAATGATGATTTATTTCA	841	
QY	540	ACGAATCTGCCAG 552		
DB	842	AAGAATCTGCCAG 854		
RESULT	4			
LOCUS	HS009178	3951 bp	mRNA	PRI
DEFINITION	Human dihydroxyimidine dehydrogenase mRNA, complete cds.			28-DEC-1994
ACCESSION	U09178			
VERSION	U09178.1	GI:558304		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
AUTHORS	1 (bases 1 to 3951)			
TITLE	Yokota,H., Fernandez-Salguero,P., Furuya,H., Lin,K., McBride,O.W., Podoshun,B., Schaeke,K.D. and Gonzalez,F.J. cDNA cloning and chromosome mapping of human dihydroxyimidine dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and congenital thymine uraciluria			
JOURNAL	J. Biol. Chem. 269 (37), 23192-23196 (1994)			
MEDLINE	94365020			
REFERENCE	2 (sites)			
AUTHORS	Eggenk,G., Engel,H., Viend,G., Terpstra,P. and Wilholt,B.			
TITLE	Rubredoxin reductase of Pseudomonas oleovorans. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints			
JOURNAL	J. Mol. Biol. 212 (1), 135-142 (1990)			
MEDLINE	90204534			
REFERENCE	3 (sites)			
AUTHORS	Porter,D.J., Chestnut,W.G., Merrill,B.M. and Spector,T.			

TITLE	Mechanism-based inactivation of dihydroxyrimidine dehydrogenase by 5-ethynyluracil
JOURNAL	J. Biol. Chem. 267 (8), 5236-5242 (1992)
MEDLINE	92184771
REFERENCE	4 (sites)
AUTHORS	Dupuis,A., Skehel,J.M. and Walker,J.E.
TITLE	A homologue of a nuclear-coded iron-sulfur protein subunit of bovine mitochondrial complex I is encoded in chloroplast genomes
JOURNAL	Biochemistry 30 (11), 2954-2960 (1991)
MEDLINE	91175743
REFERENCE	5 (sites)
AUTHORS	Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.J.
TITLE	Interaction of pyrophosphatase moieties with alpha-helices in dinucleotide binding proteins
JOURNAL	Biochemistry 24, 1346-1357 (1985)
REFERENCE	6 (bases 1 to 3951)
AUTHORS	Gonzalez,F.J.
TITLE	Direct Submission
JOURNAL	Submitted (22-APR-1994) Frank J. Gonzalez, National Cancer Institute, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES	Location/Qualifiers
source	1..3951
CDS	/organism="Homo sapiens" /db_xref="taxon:9606" /tissue_type="liver" 82..3159 /codon_start=1 /product="dihydroxyrimidine dehydrogenase" /protein_id="AA57474.1" /db_xref="GI:558305" /translation="MAPVLSKDSADIESILALNPRQTHATLCSTSAKKLDRKKMKRN PDKNCNCKLENPDIDIKHTTIGERALREARCKADAPKQSCPNTLIDKSFIT STANKNYGAAKMTESDNPGLTGWVCPSPDLCVGCNLYATEBEPINIGLQDFAT EYFKAMSIQIRNPSLPPEKSEASAKIALRGAASISCSFLARINSDITLFE KOEYVGLSTSEIPQRLPYDVNVEIEIMKDKLGVKILGKSLSVEMTLSTLEKCY KAAFIIGLPEPKNDALIFQGLTODGFTYSKDFPLVAKSGKAGMCSPPLPSIRGV VVLVAGDTAFDCQDTSALRCGARVIVFRKGVNIRAVPEMEELAKKEKCEPLPS PKVIYKGRIRYAMQVPRBEDQETGKNMEDDQVHLKADVIASFGSVLSDPKVEA LSPILRNWGLEPVEDPMTQSEAVWAFAGDVGALNTTVESVNDGKQASWY1HKYVO SOYASVSAPKPELPFTPIDVIDIVSYEMAGLKNTPFGLASATPATSTSMIRAREA GWGFALTKEFSLDKDIDVTNVSPIIRKRTSGPMWGGOSFLNIELISEKTAAYWCS VTELEKADPDNIVIASIMCSYNKNDWTELEKSEDSGADALENLSCPHMGKRGWL ACGODELVRNICRWRQAVQIDPEFAKLTPNVDIVSIRAAKREGANCVYATNYS LMGLKSDGTPWPVAVGIKATRTTGVSGLTIRPALAAVTSIRALDFPTLATGIDS AESGLQFLHSGASVLOVCSAIONDFTVEDCTGKALYLSIELEDDWQSPAT VSHQKGPVRIAEIMDKRLPSFGPYLEIMRKDIATKRIKLEONAVFPLKSCFIP KRPIPIKDVIGALQYLGTFGLSNVEGVAAIDEMKICNGKCYMTCNDSCYQAIQ FDPETHLPITVDITDCTGCTCLSVCPIDVCIKMSRTTPYEPKRGVPLSVNPGC" 1084..1134 /citation=[5] /function="catalytic cofactor NADPH/NADP binding site" misc-feature 1492..1524 /citation=[2] /function="electron transfer center, FAD binding site" misc-feature 2062..2175 /citation=[3] /function="uracil (substrate) binding site" misc-feature 2938..2973 /standard_name="iron-sulfur center" /citation=[4] /function="catalytic cofactor [4Fe-4S] binding site"
BASE COUNT	1153 a 785 c 896 g 1117 t
ORIGIN	
Query Match	20.1%; Score 172.8; DB 97; Length 3951;
Best Local Similarity	93.8%; Pred. No. 2.8e-24;
Matches 180; Conservative	0; Mismatches 12; Indels 0; Gaps 0;
QY	244 CTCAATATCTTTACTCTTTCATGAGACATTTGCAAAATGTTCCCCCAATATCATCCG 303
DB	1797 CACCAAACTTCTCTGATAGACATTTGCAAAATGTTCCCCCAATATCATCCG 1856

QY 304 GGAACCACTCTGCGCCCATGTATGCGCCCTGGACAAGCTCTTTCGAATATTGAGCT 363
|||||
Db 1857 GGAACCACTCTGCGCCCATGTATGCGCCCTGGACAAGCTCTTTCGAATATTGAGCT 1916
QY 364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTCTCACTGACTAAAGGCTGACTT 423
|||||
Db 1917 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTCTCACTGACTAAAGGCTGACTT 1976
QY 424 CCCAGACACAT 435
|||||
Db 1977 CCCAGACACAT 1988

RESULT 5
AR027740 AR027740 3957 bp DNA PAT 29-SEP-1999
LOCUS Sequence 1 from patent US 5856454.
DEFINITION AR027740
ACCESSION AR027740
VERSION AR027740.1 GI:5938560
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 3957)
AUTHORS Gonzalez, F. J. and Fernandez-Salguero, P.
TITLE cDNA for human and pig dihydropyrimidine dehydrogenase
JOURNAL Patent: US 5856454-A 1 05-JAN-1999;
FEATURES
source 1..3957
location/Qualifiers
BASE COUNT 1156 a 787 c 897 g 1117 t
ORIGIN

Query Match 20.18; Score 172.8; DB 9; Length 3957;
Best Local Similarity 93.8%; Pred. No. 2.8e-24;
Matches 180; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 244 CTCATATCTTTACTCTTTCATGAGACATGTGACAAATGTTCCCATTAATCATCG 303
|||||
Db 1803 CACCAAACTTCTCTCTGTGTAAGACATTTGACAAATTTTCCCGACATCATCCG 1862
QY 304 GGAACCACTCTGCGCCCATGTATGCGCCCTGGACAAGCTCTTTCGAATATTGAGCT 363
|||||
Db 1863 GGAACCACTCTGCGCCCATGTATGCGCCCTGGACAAGCTCTTTCGAATATTGAGCT 1922
QY 364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTCTCACTGACTAAAGGCTGACTT 423
|||||
Db 1923 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTCTCACTGACTAAAGGCTGACTT 1982
QY 424 CCCAGACACAT 435
|||||
Db 1983 CCCAGACACAT 1994

RESULT 6
AB003063 AB003063 3078 bp mRNA PRI 21-JAN-2000
LOCUS Homo sapiens mRNA for dihydropyrimidine dehydrogenase, complete
DEFINITION
ACCESSION AB003063
VERSION AB003063.1 GI:6729337
KEYWORDS dihydropyrimidine dehydrogenase.
SOURCE Homo sapiens male lymphocyte cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Ogura, K., Nishiyama, T., Takubo, H., Kato, A., Okuda, H., Arakawa, K.,
Fukushima, M., Nagayama, S., Kawaguchi, Y. and Watabe, T.,
TITLE Sulfidai inactivation of human dihydropyrimidine dehydrogenase by
(E)-5-(2-bromovinyl)uracil derived from the antiviral, sorivudine

JOURNAL Cancer Lett. 122 (1-2), 107-113 (1998)
MEDLINE 98124145
ERRATUM: [[published erratum appears in Cancer Lett 1998 Jun
19;128(2):229]]
REFERENCE 2 (bases 1 to 3078)
AUTHORS Ogura, K.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1997) to the DDBJ/EMBL/GenBank databases.
Kenichiro Ogura, Tokyo University of Pharmacy and Life Science,
Department of Drug Metabolism and Molecular Toxicology, 1432-1
Horiuchi, Hachioji, Tokyo 192-0392, Japan
(E-mail:ogurakeps.royaku.ac.jp, Tel:+81-426-76-4518,
Fax:+81-426-76-4517)

FEATURES
source 1..3078
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="lymphocyte"
/sex="male"
1..3078
/codon_start=1
/product="dihydropyrimidine dehydrogenase"
/protein_id="BA8978.1"
/db_xref="GI:6729338"
/translation="MAVLSKDSADIESILALNPRQTHATLCSYSAKKLKKHKN
PDKNFCNCKLNNFDDIKHTLGERGALREARKLCKADAPKSCPNLDIKSFIT
SINKNYYGAAMKIFSDNP.LGTCGWCPESLDLGVGMLYATEBPINIGLQOFAT
EVRKMSIPIQIRNPSLPPKMSVSAKALFAGPASISGASFLARLGYDITFE
KQRYGSLSTSPQRLPYDVNFEIETLTKMGDLYITGKSLSVNEMTLSTLKEKY
KAFTIGLPEPKNDATFQGLTDDQGYTSKDLPLVAAGSRAGMOCASPLSTIRGV
VIVAGDTAFDCATSALRCGARVFIIVRKGVNRAVEEMELAKEKCELPFLS
PRKIVYVGRIVAMQEVRTEDQETGKMNEDDMHLKADVVISAGVSLDPKYEA
LSPLKERNMGLPEVDEPMOTSEAWFAGGVYGLANTVESVNDGKOSMYIHKYVO
SOYGSVSAKPELPLFTPTIDIVDISVEMAGLKFIPGLASATPATSTSMIRAFEA
GMGFALIKRTSLDKDITVNSPRILITGTSGBPVGSGSFLNIELISKTAAYCOS
VTELRADFPNIVIASIMSYNSNDWTEIATKSEDSGDALELNLSCPHGKERMG
ACQDDPELVNIRWRAVOIPFEAKLPNVNIDVIRAAKEGANGAVTATNVSG
LMGKSDGTPWPAVGIARKRTTGVSGGTARIPALAAVSIARALPGFPIIATGIDS
AESGLQFLHSGASVLOVCSAIONODFTVEDCTGKALLYLKSLIEELDDMGSPAT
VSHQKGPVRIAEIMDKLPSFGPLLEORRTIAENKIRLEONVAFSPKRNCFIP
KRPLPTKVDYIGKALQYTGFEGLSVNEOVYAMIDEMKINGCKMYMCNDSGYQAIQ
FDPETHLPITTDCTGCTTCLSVCPYDICKAVSRTPPEKRGVPLSVNVC"

BASE COUNT 875 a 643 c 734 g 826 t
ORIGIN

Query Match 19.9%; Score 171.2; DB 85; Length 3078;
Best Local Similarity 93.2%; Pred. No. 5.8e-24;
Matches 179; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 244 CTCATATCTTTACTCTTTCATGAGACATGTGACAAATGTTCCCATTAATCATCG 303
|||||
Db 1716 CACCAAACTTCTCTCTGTGTAAGACATTTGACAAATGTTTCCCGACATCATCCG 1775
QY 304 GGAACCACTCTGCGCCCATGTATGCGCCCTGGACAAGCTCTTTCGAATATTGAGCT 363
|||||
Db 1776 GGAACCACTCTGCGCCCATGTATGCGCCCTGGACAAGCTCTTTCGAATATTGAGCT 1835
QY 364 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTCTCACTGACTAAAGGCTGACTT 423
|||||
Db 1836 CATCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTCTCACTGACTAAAGGCTGACTT 1895
QY 424 CCCAGACACAT 435
|||||
Db 1896 CCCAGACACAT 1907

RESULT 7
HSU20938 HSU20938 4409 bp mRNA PRI 04-APR-1997
LOCUS Human lymphocyte dihydropyrimidine dehydrogenase mRNA, complete
DEFINITION
ACCESSION U20938

VERSION U20938.1 GI:1926407
KEYWORDS
SOURCE human.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 4409)
AUTHORS Diasio, R.B., Beavers, T.L. and Carpenter, J.T.
TITLE Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity
JOURNAL J. Clin. Invest. 81 (1), 47-51 (1988)
MEDLINE 88087908
REFERENCE 2 (bases 1 to 4409)
AUTHORS Lu, Z.H., Zhang, R. and Diasio, R.B.
TITLE Purification and characterization of dihydropyrimidine dehydrogenase from human liver
J. Biol. Chem. 267 (24), 17102-17109 (1992)
MEDLINE 92381021
REFERENCE 3 (bases 1 to 4409)
AUTHORS Johnson, M.R., Albin, N., Shabinian, H. and Diasio, R.B.
TITLE Identification of a frameshift in the DNA coding for dihydropyrimidine dehydrogenase (DPD) in a DPD deficient patient exhibiting 5-fluorouracil toxicity
Unpublished
4 (bases 1 to 4409)
AUTHORS Johnson, M.R.
TITLE Direct Submission
Submitted (10-FEB-1995) Pharmacology, University of Alabama at Birmingham, 1670 University Blvd., Birmingham, AL 35294-0019, USA
On Apr 4, 1997 this sequence version replaced gi:693912.
COMMENT
FEATURES
source
1..4409
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="Lymphocyte"
102..3179
/EC_number="1.2.1.2"
/note="DPD: dihydropyrimidin dehydrogenase"
/codon_start=-1
/product="dihydropyrimidine dehydrogenase"
/protein_id="AAB51366.1"
/db_xref="GI:693912"
/translation="MAPVLSKDSADIESILALNPRQTHATLCTSTAKKIDKKHMKRN
PDKNCNCELENNFDIKHTTIGERALKREARCLKCAPQKSCPTNLDIKSFIT
STANKNYGAKAKIFSDNPLGLTCGMVCPISDLGCGCNLYATEEGINIGLQOYAT
EVKKAMSIPOIRNPSLPPEPKMEASAKTALGAGPASISCAFLARLGYSDITTFE
KOEYVGLSTSEIPORLPYDVNFELEMLKDLGVNITICGSLSVNEMTLSTKERGY
KAPFVIGLPEPKMDAIFQGLTODQGYTSKDFLPLVAKSGKAMCACHPLPSIKGY
VIVLAGDITAFDCATSLRCGARVFLVERKGVNIRAVPEEMELAKEECFELPLFS
PRKVIYGGRIVAMOFYRTQDETKGWNEDQMHILKADYVTSARGSVLSDPKVEA
LSPIKRRMGLPEVDEPTNOTSEAWPFAGDVGANTYTESVNDGKQASWYTHKYYO
SOTGASVASAPELPLFYTPDLDIVISEMAGLKFINFGLASATPATISTMIRAREA
GWFALKTSELDKDIVTNVSPIRIGTTSGPYGGQSSFLNILEISETTAWMOS
VTELKADFPNIIYIASIMCSYNNMDWTELAKESEDGADALENLISCPHMGGRGKL
ACQDEPLVNICIRWVROAVRIPEFAKLTNPVNTDIVISIAAAREGGANGVATNTVSG
LMELKSDGPMVPAVGIAKRTTGVGSGTAIRTAIRAVTISIAALGEPFLTAGGIDS
AESGLQFLHGSAGVLOVCSAIONQDFTYEDTCTGKALILYLSIELODMDQSPAT
VSHQKGPVRIAEIMDKLPSFGPYLEKCKKIIAEKRIKLEONAVAFSLKNCNCFIP
KRPIPTIKDVIKALQYLGTFELSNVEQVAMIDEKMCINCKRCYMTKNDSCGYOIO
FDEPFLPTTDECTCTCTGATAGGACATTGTGCAAAAGTTTCCCCAGAAATCATCCG 1876

BASE COUNT

1317 a 845 c 949 g 1298 t

ORIGIN

Query Match 19.9%; Score 171.2; DB 97; Length 4409;
Best Local Similarity 93.2%; Pred. No. 5.8e-24;
Matches 179; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 244 CTCATATCTTACTCTTTCATGAGACATTGTGACAAATTTTCCCCCATATCATCCG 303
DB 1817 CACCAAACTTCTCTCTGATAGGACATTGTGCAAAAGTTTCCCCAGAAATCATCCG 1876

QY 304 GGAACACACTCTGAGCCCATGTATGGCCCTGGACAAAGCTCTTCTGATATTGACT 363
DB 1877 GGAACACACTCTGAGCCCATGTATGGCCCTGGACAAAGCTCTTCTGATATTGACT 1936
QY 364 CATCAGTGAAGAAACGGCTGCATATTGGTGTCAAAAGTGCACAGTAAGGCTGACT 423
DB 1937 CATCAGTGAAGAAACGGCTGCATATTGGTGTCAAAAGTGCACAGTAAGGCTGACT 1996
QY 424 CCCAGACACAT 435
DB 1997 TCCAGACACAT 2008
RESULT 8
BTU20981
LOCUS 4421 bp mRNA MAM 17-JAN-1997
DEFINITION Bos taurus liver dihydropyrimidine dehydrogenase mRNA, complete cds.
ACCESSION U20981
VERSION U20981.1 GI:677950
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Catartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
1 (bases 1 to 4421)
AUTHORS Lu, Z., Zhang, R. and Diasio, R.B.
TITLE Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients, and clinical implication in 5-fluorouracil chemotherapy
Cancer Res. 53 (22), 5433-5438 (1993)
MEDLINE 94036837
REFERENCE 2 (bases 1 to 4421)
AUTHORS Albin, N., Johnson, M.R. and Diasio, R.B.
TITLE DNA cloning of bovine liver dihydropyrimidine dehydrogenase
DNA Seq. 6 (4), 243-250 (1996)
MEDLINE 97069948
REFERENCE 3 (bases 1 to 4421)
AUTHORS Johnson, M.R.
TITLE Direct Submission
Submitted (13-FEB-1995) Pharmacology/Toxicology, University of Alabama at Birmingham, 1670 University Blvd., Birmingham, AL 35294-0019, USA
FEATURES
source
1..4421
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="liver"
75..3152
/EC_number="1.3.1.2"
/note="DPD: dihydropyrimidine dehydrogenase; DPD"
/codon_start=-1
/product="dihydropyrimidine dehydrogenase"
/protein_id="AAB40985.1"
/db_xref="GI:677951"
/translation="MAPVLSKDVADIESILALNPRQSRATLSTLAKKIDKKHMKRN
PDKNCNCELENNFDIKHTTIGERALKREARCLKCAPQKSCPTNLDIKSFIT
STANKNYGAKAKIFSDNPLGLTCGMVCPISDLGCGCNLYATEEGINIGLQOYAT
EVKKAMSIPOIRNPSLPPEPKMEASAKTALGAGPASISCAFLARLGYSDITTFE
KOEYVGLSTSEIPORLPYDVNFELEMLKDLGVNITICGSLSVNEMTLSTKERGY
KAPFVIGLPEPKMDAIFQGLTODQGYTSKDFLPLVAKSGKAMCACHPLPSIKGY
VIVLAGDITAFDCATSLRCGARVFLVERKGVNIRAVPEEMELAKEECFELPLFS
PRKVIYGGRIVAMOFYRTQDETKGWNEDQMHILKADYVTSARGSVLSDPKVEA
LSPIKRRMGLPEVDEPTNOTSEAWPFAGDVGANTYTESVNDGKQASWYTHKYYO
SOTGASVASAPELPLFYTPDLDIVISEMAGLKFINFGLASATPATISTMIRAREA
GWFALKTSELDKDIVTNVSPIRIGTTSGPYGGQSSFLNILEISETTAWMOS
VTELKADFPNIIYIASIMCSYNNMDWTELAKESEDGADALENLISCPHMGGRGKL
ACQDEPLVNICIRWVROAVRIPEFAKLTNPVNTDIVISIAAAREGGANGVATNTVSG
LMELKSDGPMVPAVGIAKRTTGVGSGTAIRTAIRAVTISIAALGEPFLTAGGIDS
AESGLQFLHGSAGVLOVCSAIONQDFTYEDTCTGKALILYLSIELODMDQSPAT
VSHQKGPVRIAEIMDKLPSFGPYLEKCKKIIAEKRIKLEONAVAFSLKNCNCFIP
KRPIPTIKDVIKALQYLGTFELSNVEQVAMIDEKMCINCKRCYMTKNDSCGYOIO
FDEPFLPTTDECTCTCTGATAGGACATTGTGCAAAAGTTTCCCCAGAAATCATCCG 1876

	KKLPISVKDVGKALQYIGTYGELNNTEQVAVAVIDEEMCIKCGACYMTCNDSPGOA FDPETHLPTVTDCMCTGCCTLCISVCPDIIDCIMVSTPPEKRRLPLAVNVVS"
BASE COUNT	1315 a 874 c 958 g 1274 t
ORIGIN	
Query Match	18.0%; Score 154.8; DB 7; Length 4421;
Best Local Similarity	88.4%; Pred. No. 9.1e-21;
Matches 168; Conservative	0; Mismatches 22; Indels 0; Gaps 0;
QY	246 CAATATTCTTACTCTTTTCATGAGAGCATGTGACAATGTTTTCCCATTAATCATCCGGG 305
Db	1792 CCAAAACTTTCTCTCTTGATAGAATAGTACAAATGTTTCACCCAGAACATCCGGG 1851
QY	306 GAACCACCTCTGGCCCCCATGTATGGCCCTGGACAAAAGCTCTTTCTGAATTATGAGCTCA 365
Db	1852 GGACCACTCTGGCCCCCATGTATGGCCCTGGACAAAAGCTCTTTCTGAATTATGAGCTCA 1911
QY	366 TCAGTAGAAAAAGGCTGCATATTGTTGCCAAGTGTCACAGCAACCAAGCTGACTCC 425
Db	1912 TCAGTAGAAAAAGGCTGCATATTGTTGCCAAGTGTCACAGCAACCAAGCTGACTCC 1971
QY	426 CAGACACAT 435
Db	1972 CAGCACATAT 1981
RESULT	9
SSU09179	
LOCUS	SSU09179 4441 bp mRNA MAN 28-DEC-1994
DEFINITION	Sus scrofa dihydropyrimidine dehydrogenase mRNA, complete cds.
ACCESSION	U09179
VERSION	U09179.1 GI:558306
KEYWORDS	.
SOURCE	Pig.
ORGANISM	Sus scrofa
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus. 1 (bases 1 to 4441)
AUTHORS	Yokota,H., Fernandez-Salguero,P., Furuya,H., Lin,K., McBride,O.W., Rodeschun,B., Schnackerz,K.D. and Gonzalez,F.J.
TITLE	cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and congenital thymine uraciluria J. Biol. Chem. 269 (37), 23192-23196 (1994) 94365020
JOURNAL	
MEDLINE	
REFERENCE	
AUTHORS	Engink,G., Engel,H., Vriend,G., Terpstra,P. and Witholt,B.
TITLE	Rubredoxin reductase of Pseudomonas oleovorans. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints J. Mol. Biol. 212 (1), 135-142 (1990) 90204534
JOURNAL	
MEDLINE	
REFERENCE	
AUTHORS	Porter,D.J., Chestnut,W.G., Merrill,B.M. and Spector,T.
TITLE	Mechanism-based inactivation of dihydropyrimidine dehydrogenase by 5-ethynyluracil J. Biol. Chem. 267 (8), 5236-5242 (1992) 92184771
JOURNAL	
MEDLINE	
REFERENCE	
AUTHORS	Dupuis,A., Skehel,J.M. and Walker,J.E.
TITLE	A homologue of a nuclear-coded iron-sulfur protein subunit of bovine mitochondrial complex I is encoded in chloroplast genomes Biochemistry 30 (11), 2954-2960 (1991) 9115743
JOURNAL	
MEDLINE	
REFERENCE	
AUTHORS	Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.J.
TITLE	Interaction of pyrophosphatase moieties with alpha-helices in dinucleotide binding proteins Biochemistry 24, 1346-1357 (1985) 6 (bases 1 to 4441) Gonzalez,F.
JOURNAL	
MEDLINE	
REFERENCE	
AUTHORS	Gonzalez,F.
TITLE	Direct Submission

[illegible]

ACCESSION	AR027741		GI:5938561
VERSION	AR027741.1		
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 4447)		
TITLE	Gonzalez,F.J. and Fernandez-Salguero,P.		
JOURNAL	cDNA for human and pig dihydropyrimidine dehydrogenase		
FEATURES	Patent: US 5856454-A 3 05-JAN-1999;		
source	Location/Qualifiers		
	1..4447		
BASE COUNT	/organism="unknown"		
ORIGIN	1283 a 930 c 1000 g 1234 t		
Query Match	17.6%;	Score 151.6;	DB 9; Length 4447;
Best Local Similarity	87.4%;	Pred. No. 3.8e-20;	
Matches 166;	Conservative 0;	Mismatches 24;	Indels 0; Gaps 0;
OY 246	CAATATCTTACTCTTCGATGAGGACATTGTGCACAAATGTTCCCATTAATCATCGGG 305		
Dd 1805	CCAATACTTCTCCTCTGATAGAAGACATAGACAAATGTCTCACCAAGATCGCCGG 1864		
OY 306	GAACCACCTCTGGCCCCATGTATGGCCCTGGACAAGCTCCTTTTGATATATTGACTCA 365		
Dd 1865	GGACTACTCTGGCCCCATGTATGGCCCTGGACAAGCTCCTTTTGATATATTGACTCA 1924		
OY 366	TCAGTAGAAGAAAAGCGCTGCATATTGGTGTCAAAGTGCATGAACTAAAGGCTGACTTC 425		
Dd 1925	TCAGTAGAAGAAAAGCGCTGCATATTGGTGTCAAAGTGCATGAACTAAAGGCTGACTTC 1984		
OY 426	CAGACACAT 435		
Dd 1985	CAGACATAT 1994		
RESULT 11			
D85035			
LOCUS	D85035 4358 bp mRNA	ROD	19-SEP-1998
DEFINITION	Rattus norvegicus mRNA for dihydropyrimidine dehydrogenase,		
	complete cds.		
ACCESSION	D85035		
VERSION	D85035.1 GI:3628592		
KEYWORDS	DDP; dihydropyrimidine dehydrogenase.		
SOURCE	Rattus norvegicus (Strain:Mistar) 7-d-old male liver cdna to mRNA,		
	clone_lfp:lambda ZAPII CDNA.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 4358)		
AUTHORS	Kimura,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAY-1996) to the DDBJ/EMBL/GenBank databases.		
	Masahito Kimura, Kobe Gakuin University, Faculty of Nutrition; 518		
	Aisise, Ikawadani-cho, Nishi-Ku, Kobe, Hyogo 651-21, Japan		
	(Tel.:078-974-1551(ex.3242), Fax:078-974-5569)		
REFERENCE	2 (bases 1 to 4358)		
AUTHORS	Kimura,M., Sakata,S.F., Matoba,Y., Matsuda,K., Kontani,Y.,		
	Kaneho,M. and Yamaki,N.		
	Cloning of rat dihydropyrimidine dehydrogenase and correlation of		
	its mRNA increase in the rat liver with age		
	J. Nutr. Sci. Vitaminol. 44 (4), 537-546 (1998)		
FEATURES	99037149		
source	Location/Qualifiers		
	1..4358		
	/organism="Rattus norvegicus"		
	/strain="Mistar"		
	/db_xref="taxon:10116"		
	/clone_lfp="lambda ZAPII CDNA"		
	/dev_stage="7-d-old"		

```

gene
CDS
    /sex="male"
    /tissue_type="liver"
    59..3136
    /gene="Ddp"
    59..3136
    /gene="Dpd"
    /EC_number="1.3.1.2"
    /codon_start=1
    /product="O6-methylguanine dehydrogenase"
    /protein_id="BAA33218.1"
    /db_xref="GI:3628593"
    /translation="MAGVLSRDADPDIESTIALNPRIOAHATLRSTMAKILDKKHKMKRR
    TDKRCFCEKLENNEDDIKHTTLCGERALRAVAQCICADAPCKSCPTSLDICKSFTI
    STAKNYYGAKLIFSDNPGLITGMCGPTSDLCVGGONLHATEEGLPNIGLDOFAT
    EEFKANNI PQIRSPLLPPEHEMPEAYSAKIALFGAPASISCAFLARLGSDITIEE
    KOEAVGGLSTSELPEQFRLPYDVVFELEIMLMDLVCIKIGCSISTDEMTLSTLENGE
    KAAEIGIGLEPEPKKDHIFQGLTYQVGYFTSGDPLVAKGSPPMCACSPKPLSVRGK
    VYVAGDADTIVDCATSAIRGCARVRPIYFRKGFANITVAPEMLALAEKECEFLPLPS
    PRAKIVDGTAKVGMQPYRTQDDETGNNVDEEQVYRLKADVISPPSGVLYDDPIYTER
    LSPLEKRWGLPEVNPETMOTSEPVNFGGVVMANTVYESVNDGKOASVTHEYIYO
    AOGVALPSQPLLEFLYTPDLVDISEMAGLRPNPFGLASAPASTPMIRAFEE
    GSGALTKTESLSDKDIYTNVSPRIIRGTTSGPLPQSGSFLINELISEKTAAWCHS
    VTLEKADPEPNILIASISYNNKNDMELSKMAPASGADALELNSCPHMGEEGMGI
    ACGDDPELVNIIICWVRSVRSVPREKATLPVVDIVSIAARKEGSGADGYATMTVSGG
    LMELKADGSPRPSVSGSKRTTYGVSSTTRPIALRAVTAARLRPPIIATGCTGDS
    AESLOFLHSGASVLYQVCSAIONDFVIEDYCTGLALALYLSIEELSDMDGSPPT
    MSHQGRFVPHIAELMOQKLPSEGFYLERKKILAAKIRENDONRACSPIDQRHFNFS
    OKRPAIKVDIGKSLQELGTGEINIMEQVVALIDEMCINCSCYMTCNDSGYOAIQID
    FDPETHLPYSDPCTGCTGLCSVCPIMDCIKMSRAPIYEPKRLPLAVKRYVC"

BASE COUNT      1234 a      930 c      1022 g      1172 t

ORIGIN

Query Match      17.0%; Score 146.8; DB 94; Length 4358;
Best Local Similarity 85.8%; Pred. No. 3.3e-19;
Matches 163; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 246 CAATATCTTACTCTTTCATGAGGACATTTGACAAATGTTTCCCCATATCATCCGGG 305
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1776 CCAAAACTTCTCTCTTGATGAAGGACATCTGACAAAGCTGTCACCCGAAATCATCCGAG 1835

QY 306 GAACCACTCTGGCCCGCATGTATGGCCCTGGACAAAGCTCCCTTGATGATATGAGCTCA 365
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1836 GGACCACTCTGGCCCGCTTGATGGCCCTGGACAAAGCTCTTTCTTCAACATTTGAGCTCA 1895

QY 366 TCAGTGAGAAAGCGCTGCATTTGGTGTCAAGAAGTGCATGAACATAAGGCTGACTTCC 425
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1896 TCAGTGAGAAAGCAAGCTGCATTTGGTGTTCACAGTGTACCGGAACATAAGGCTGACTTCC 1955

QY 426 CAGACACACAT 435
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1956 CGGACACACAT 1965

RESULT 12
CELC25F6      43803 bp      DNA      INV      02-NOV-1995
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 43803)
Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kersey,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurry,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,T., Roopra,A.,

```

TITLE
elegans 368 (6466), 32-38 (1994)

JOURNAL
MEDLINE
94150718

REFERENCE
2 (bases 1 to 43803)

AUTHORS
Bentley, D.

TITLE
The sequence of C. elegans cosmid C25F6

JOURNAL
3 (bases 1 to 43803)

AUTHORS
Waterston, R.

TITLE
Direct Submission

JOURNAL
Submitted (30-OCT-1995) Robert Waterston

COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1HQ, England
e-mail: r.waterston@wustl.edu and jess@sanger.ac.uk

NEIGHBORING COSMID INFORMATION:
The 5' cosmid is C41A3, 1000 bp overlap; 3' end lies in a gap
followed by the cosmid J23F2. Actual start of this cosmid is at
base position 1 of CELC25F6; actual end is at base position 43803
of CELC25F6.

NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source
1. 43803
Location/Qualifiers
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/map="x"
complement(1397..1495)
/note="Leu: codon recognized: CTA"
complement(5095..5656)
/gene="C25F6.6"
complement(join(5095..5215,5259..5385,5443..5535,
5584..5656))
/gene="C25F6.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AA80431.1"
/db_xref="GI:1049456"
/translation="MMSRFLSKILIVSCLIGLVADASSMYARRRTASERLMSII
RPMFONPERSSQIKQISLIQINPNRALPCCKDESGSICKNLRRTDKLFTKQCTE
PDFSLVCCSSQASIGISYRRKQKOVFIITLVI"
complement(6056..6128)
/note="Lys: codon recognized: AAA"
16243..20895
/gene="C25F6.4"
join(16243..16401,16449..16601,17110..17257,17309..17421,
17576..17675,17720..17825,18125..18202,18252..18358,
18406..18572,18930..19243,19293..19592,20318..20513,
20569..20718,20773..20895)
/gene="C25F6.4"
/note="similar to protein tyrosine kinase"
/evidence=not_experimental
/protein_id="AA80432.1"
/db_xref="GI:1049457"
/translation="MQGDIADSOITASSRDKOSVGPQNAKLHSELASGWCCKPQI
NSKSYFLQVLTNDTFLITSVEYGRGNGRGFASHYMDVLRPSQWLRVNRGTS
HYWDGNEFTTTPVIRVLDPIYASRIKRPVSSNTRVNCRAEIHGCKHGVYVST
VPDGRSLDTLDFKDSMFEDSQITYESGIKRGSWPNARRFSASPEKKNMNSMIG
MNRDTDGRITLILFEFEVHNFTDVLATFNGRIDGIDVIRPSQDKTPELPISOSSSE
ROSINSTRSDYDFPLANRAGRKRVRISIKSSDMFLTEVHFTSAANLTLISKIKP
PSSAAHQQLLYVCGIITFLTFACVAYCYSVCLAKRQKKSVDSNVKKDLITTHGNKP
TCHVFPSSNGKLNGHYEVANDILYARSOKSTLLSVSSKSTFSCRAIDPTWDFNPPPP

gene
CDS
join(22629..22778,22996..23109,23156..23244,23886..24135,
24185..24653,24703..24875,24923..25050,25114..25196,
25344..25495,25544..26161,26225..26956,27003..27136,
27190..27352)
/gene="C25F6.3"
/note="coded for by C. elegans cDNA yk28e9.3; coded for by
C. elegans cDNA yk28e9.5; coded for by C. elegans cDNA
yk28h9.3; coded for by C. elegans cDNA yk28h9.5; coded for
by C. elegans cDNA yk40e4.5; coded for by C. elegans cDNA
yk5h3.5; weakly similar to dihydroorotate dehydrogenase
and glutamate synthase"
/codon_start=1
/protein_id="AA80433.1"
/db_xref="GI:1049458"
/translation="MSYDFAFCSTIFRCKFNIIIFTAMPEKPTNTSPNNILSLK
DSPIESLILINPKVQDRANVPSAVTKRKKHMKRKEEGCGSTGSESKLDFDI
KHTLISERGALEKEMRCLKADAPCQSCPTQDVSVFTISNKNYGAARQILSDN
PLGLTCMIGPTSDLCVGCNLSQSEGAINIGLOQYACDVKRQNNVOIYSEKRE
NRNASHKEOVALLICGCPASISCAFIARLCTPTIYEKRAYIGLSAEIDPFLPY
DVDFEIQIADIGVIGETNRPJGKDLTLAKLKEGAAVAAPVIGICINPEKPIDLEG
LITENGFYTSKNRPLPAVAASKPGMCKRTPITMGRKRVVYAGADTMDCAATLR
CGASRVITAFKRGFTGIRAYDEMEAKKEKCEFLPISARAKINVDGRIVSTEFKT
EQDNGKWEDEQVILKCDYVISAAGSTLKEADVLSALQPCQLKKGWIEVDSTQ
QTSERWAFAGDVGAVAEVTYESVNDKRIAMMMHRTIOSLHONVSELPEDPFTP
IDEVDISVDMGVKFEENPFGIASAPPTSPCKRRAREQCMGILTKTQGLDGLTVN
VSPRIVRGSTGPLYGNOGSGFNIELISKSECEYMLQCRELRKDPHPIYASIMG
VYNKADMIETALKESEAGADITELNLSCPHMEKKGMLACGOSPEITVEICWAVAC
VKIPEPKPTNITDVREIARAADGASCVTATNTVSSLMHKKACGNMARPAGSKR
TTGMSGSATRPITAMKAVSIALNELGFTIMATGIESLETGLFMGASVLYQCS
AVQMDQFTVVDVCTGLKALLYLISGASLNMWQGSPIPERKHGKRPILQOGKKMF
FGKRYDEREKELEAIKLSSENLIDTENTHFSRPTQVSRVPTVEDYIGALPRIGPV
TLNQDEOKVAIIDDMCINCCKCYMTNDGSOYALTFDPYHDPHTEDEDCCTICY
SVCPIDPECIEMVPRTPGWKAPKRGKRSVETGPYKVKVQDGRVILDTTGMQ"
34301..39725
/gene="C25F6.2"
join(34301..34613,35496..35812,35889..36107,36155..36251,
36299..36389,36443..36794,36851..37157,37216..37684,
37743..38266,38350..38632,38680..38801,38857..38870,
38931..38949,39228..39253,39483..39725)
/gene="C25F6.2"
/note="coded for by C. elegans cDNA yk25e5.3; coded for by
C. elegans cDNA yk25e5.5; similar to repeat guanylate
kinase domain of D. melanogaster lethal(1) discs large-1
tumor suppressor protein (SP:DLG1, DROME, P31007) and R.
norvegicus postsynaptic density protein 95 (PSD-95)
(SP:PSD9-RAT, P31016)"
/codon_start=1
/protein_id="AA80434.1"
/db_xref="GI:1049459"
/translation="MAACLCYVCVACGCKRFLPTLSDKLCILIAOFFLEVENASFLP
SSVRLLEALKPVYEFYEOSOKIYVDQENFNPORPASKERSBWRLEPESMSHES
SEKAKRAIENVEDYCOITLRHGBELATNLERIYTPKSLHMSLIDTIDLYEOTLS
ERKSDAKRNEVRKRVIRLGGPHSTISRPAATTSNTNLSTPLISDLDRGFS
YLNGGGLNGNGGLNGLLSPYSSSTYLRHROGRTSHDQTEETTRTVDPG
LERRVHEHTGVYIDHGRKMELENTVLEKGTGLGFSITGMDQPTEDGDTSYIVNI
IEGGAALADGRMKRNDITTAANNNTNCEVAKHEVAVNLKSGANVSLSLKRRDEALP
IGNREGSTSYLRSGVPSVAGNLQAHNSPAPLHPRPVPHVHSLRSLQSYGR
SPRNPVSVIDLYGKARGSLGSGGQNEVKKDPTIYVYKIIIEGAALDLDRRGD
KILEVDHSLINTTTHENAVNLAKTGRVLLIQOGTAIFINDSASQOFTPTPLRP
SSVDIYRSQSSSHLSYSGPLNTSSVSQALPAIDLEPPVQLVAGQGLDFNVLG
EDNPEIYISVLPDGVADLSGNVKTGVLELVAVYLRATNHEAALRNAGNPPYL
TLQYRPOEYIOIFESKIEKLNDYIAOSRMQTLSEKSEYVATLDPYSENSVAPRS
MGFYAGIILHINSDEEMWTAKRVHNEETGELVYPSKREKREKRRRVQNS
GSGGLGNNSTGTGLENRGRSOLSPSRKRPYKSPMDRLNDINPSSNVAEPWASYO
AVEQQAIVYRPVYIIGALKDRIINDELVANDPSKFSYCCPHHSTRPREEVAGRDYV
VNKRHMEDEVKNNLFEAGFQNNLTGTSIOSVRYVANOGRHICILDVSGNAIRLOS

CDS complement(join(40144..40361,40432..40549,41242..41250))

Query Match 8.3% Score 71.4; DB 4; Length 303092;

Best Local Similarity 63.9%; Pred. No. 0.00016; Mismatches 61; Indels 0; Gaps 0;

Matches 106; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 267 AGGACATTGTGACAAATGTTTCCCATATATCATCCGGGACCACTCTGGCCCATGT 326

Db 165897 AGGATCTGTGTCAGCAATGTCTCGCGGCATCTGAGGGGACCACTGCGGCTACAAAGT 165898

QY 327 ATGGCCCTGGACAAAGCTCCTTCTGAATATGAGCTCATCATGAGAAACGGCTGCAT 386

Db 165837 ATGGACCGCAGCAGGCGTGTCTTCTGAACTATCTCGAAGAGAGCGGCCGAGT 165778

QY 387 ATGGTGTCAAGTGTCTACCTAAGAGGCTGACTTCCGACAGACAGT 435

Db 165777 ACTGGCTGAATTCGATTTGAGAGACTGAAGCGTGTCTCCGAGAGAGT 165729

RESULT 15

DMU65491

LOCUS DMU65491 2348 bp mRNA INV 26-SEP-1996

DEFINITION Drosophila melanogaster Dreg-3 protein mRNA, complete cds.

ACCESSION U65491 GI:1561729

VERSION U65491.1 GI:1561729

KEYWORDS

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 2348)

Van Gelder, R.N., Bee, H., Palazzolo, M.J. and Krasnow, M.A.

Extent and character of circadian gene expression in Drosophila

melanogaster: identification of twenty oscillating mRNAs in the fly

head

JOURNAL Curr. Biol. 5 (12), 1424-1436 (1995)

MEDLINE 96362140

REFERENCE 2 (bases 1 to 2348)

AUTHORS Van Gelder, R.N., Bee, H., Palazzolo, M.J. and Krasnow, M.A.

TITLE Direct Submission

JOURNAL Submitted (29-JUL-1996) Ophthalmology and Visual Sciences,

Washington University School of Medicine, 660 S. Euclid Avenue, Box

8096, St. Louis, MO 63110, USA

FEATURES

Source location/Qualifiers

1..2348

/organism="Drosophila melanogaster"

/strain="Canton-S"

/db_xref="taxon:7227"

358..2238

/note="similar to dihydropyrimidine dehydrogenase;

rhythmically expressed gene 3"

/product="Dreg-3 protein"

/codon_start=1

/protein_id="AAC47288.1"

/db_xref="GI:1561730"

/translation="MECCRTQENNDENDEWDEEOTRIKANPVIASRGSLIEDQYKA

ALAPLOPREGELPYVDRTYMOSSVYKQVFLGDLGAVATTVESYNDGKRAWSTHCOIQ

GLPLDTPALPLFTYDIDAVISVEMGIRFENPFGIASAPPTSTAMIRAEQGWG

FVATKTEGLDKDLVTNVSPIRVGTTSGYVPOQGLFNLIELSEKRAEYWLKISGE

LKRDPEKIVIASIMCSFNEDEWTELAKEOGADALENLISCPHOMGREGMLAG

ODELVEOISRWRKAVKLPFFIKTPNITDIYISIAAAKREPMRSAINTVGLMGL

KADSTAMPATGCKEORTYGVSGNATRPALKAISDIANRPGPILIGTIGDSGEYA

LQTHAGATYLOJCCSYQNDFTYIEDYCTALKALYLKIRRHQSNVPSGMASHHPR

SIRQAKCPEDRGRKATLGFPGYQQRDKMAELRSQKALSDAEOVATPPASNG

AAPPARIKVIDIAALDKISYKRLKIDDDMCINCCKAYMTCADSGYOAI

EFDKDIHPIVNDCTGCTLCVSCPIIDITMPPKKIPHVIRKVEEKIFYTHALSO

CQ"

BASE COUNT 538 a 649 c 684 g 477 t

ORIGIN

Query Match 8.1%; Score 69.8; DB 6; Length 2348;

Best Local Similarity 63.3%; Pred. No. 0.00034; Matches 107; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 267 AGGACATTGTGACAAATGTTTCCCATATATCATCCGGGACCACTCTGGCCCATGT 326

Db 869 AGGATCTGTGTCAGCAATGTCTCGCGGCATCTGAGGGGACCACTGCGGCTACAAAGT 928

QY 327 ATGGCCCTGGACAAAGCTCCTTCTGAATATGAGCTCATCATGAGAAACGGCTGCAT 386

Db 929 ATGGACCGCAGCAGGCGTGTCTTCTGAACTATCTCGAAGAGAGCGGCCGAGT 988

QY 387 ATGGTGTCAAGTGTCTACCTAAGAGGCTGACTTCCGACAGACAGT 435

Db 989 ACTGGCTGAATTCGATTTGAGAGACTGAAGCGTGTCTCCGAGAGAGT 1037

Search completed: November 20, 2001, 16:51:17
Job time: 6131 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 15:32:06 ; Search time 124.09 Seconds
(without alignments)
4356.703 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TCTTATGAAGATAATAATT.....ACTGGGAATAATTATTATAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	856.2	99.4	861	18	AAT91798
2	172.8	20.1	3951	20	AA222903
3	171.2	19.9	3957	17	AAT14077
4	171.2	19.9	4368	16	AAT03133
5	171.2	19.9	4369	16	AAT03143
6	154.8	18.0	4414	16	AAT03132
7	154.8	18.0	4430	16	AAT03142
8	152.6	17.7	936	22	AAF58252
9	152.6	17.7	936	22	AAF58254
10	152.6	17.7	936	22	AAF58257
11	152.6	17.7	936	22	AAF58259

12	152.6	17.7	936	22	AAF58262	Oligonucleotide D2
13	152.6	17.7	938	22	AAF58255	Oligonucleotide D1
14	151.6	17.6	4447	17	AAT14078	Pig dihydropyrimid
15	148.6	17.3	936	22	AAF58252	Oligonucleotide D1
16	148.6	17.3	936	22	AAF58254	Oligonucleotide D1
17	148.6	17.3	936	22	AAF58257	Oligonucleotide D1
18	148.6	17.3	936	22	AAF58259	Oligonucleotide D2
19	148.6	17.3	936	22	AAF58262	Oligonucleotide D2
20	148.6	17.3	938	22	AAF58255	Oligonucleotide D1
21	61.2	7.1	244	22	AAF58238	Oligonucleotide D1
22	59.8	6.9	244	22	AAF58238	Oligonucleotide D1
23	45	5.2	5852	12	AAQ11710	Dictyostelium plas
24	42.4	4.9	3648	22	AAQ11712	Plasmid PHINK181 c
25	42.4	4.9	3648	22	AAQ11713	Plasmid PHINK184 c
26	42.4	4.9	11461	22	AAQ11715	Plasmid pZU634 for
27	42.4	4.9	12766	22	AAQ11714	Plasmid pZU623 for
28	42.4	4.9	13737	21	AAA54212	Transformation vec
29	42.2	4.9	2400	18	AAT49364	Nicotiana plumbagi
30	42.2	4.9	8318	20	AAQ20264	Borrelia burgdorfe
31	41.4	4.8	745	20	AAQ39751	Gastric cancer ass
32	41	4.8	3947	19	AAV69429	Ehrlichia sp. exte
33	41	4.8	3947	21	AAZ89994	Ehrlichia antigen
34	40.2	4.7	2313	21	AAQ76449	Human ORFX ORF2004
35	40	4.6	37808	20	AAQ2780	Vector pMWX-BG DNA
36	39.6	4.6	81369	21	AAA97997	Human T gene DNA
37	39.4	4.6	639	21	AAZ98055	Human secreted pro
38	39.4	4.6	1717	20	AAQ20311	Borrelia burgdorfe
39	39	4.5	14752	20	AAQ20256	Borrelia burgdorfe
40	39	4.5	116277	20	AAQ20249	Borrelia burgdorfe
41	39	4.5	910715	20	AAQ20248	Borrelia burgdorfe
42	38.8	4.5	5810	18	AAT91323	Arabidopsis thalia
43	38.8	4.5	8911	20	AAQ13588	Enterococcus faeca
44	38.6	4.5	1400	20	AAQ27735	Human DNA marker c
45	38.6	4.5	1591	21	AAQ95204	Human UGT2B7 exon

ALIGNMENTS

RESULT 1
AAT91798
ID AAT91798 standard; DNA; 861 BP.
XX
AC AAT91798;
XX
DT 14-APR-1998 (first entry)
XX
DE Dihydropyrimidine dehydrogenase genomic fragment.
XX
KW Human; dihydropyrimidine dehydrogenase; DPD; slicing defect;
KW detection; 5-fluorouracil; cancer; anticancer; uraciluria; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 269..433
FT /*tag= a
FT /note= "Encodes amino acids 581-635 of the DPD protein"
XX
PN WO9735034-A1.
XX
PD 25-SEP-1997.
XX
PF 19-MAR-1997; 97WO-US04269.
XX
PR 20-MAR-1996; 96US-0013835.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fernandez-Salguero P, Gonzalez FJ;
XX WPI; 1997-480236/44.
DR P-PSDB; AAW30092.

XX Detecting a splicing defect in the di:hydro:pyrimidine dehydrogenase
PT gene - used to identify subjects sensitive to 5-fluorouracil, toxic
PT to individuals with DPD defects
XX
PS Disclosure: Fig 1; 38pp; English.
XX
CC A novel method has been developed for detecting the splicing defect in
CC the dihydro:pyrimidine dehydrogenase (DPD) gene. The method comprises
CC determining whether genomic DNA containing the DPD gene has a wild-type
CC intron-exon boundary for an exon that encodes amino acids (aa) 581-635
CC of the corresponding DPD protein. The present sequence represents a
CC DPD genomic fragment which encodes the amino acids 581-635. The method
CC is used specifically to determine sensitivity of subjects to the
CC anticancer agent 5-fluorouracil, which is dangerously toxic to those
CC with DPD defects. It can also be used to diagnose DPD-deficiency
CC disorders such as uraciluria.
XX
SQ Sequence 861 BP; 300 A; 140 C; 161 G; 260 T; 0 other;

Query Match 99.4%; Score 856.2; DB 18; Length 861;
Best Local Similarity 99.7%; Pred. No. 1.2e-181;
Matches 858; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTAAATGAAGATAAATATTTTGTCTTTTTCGCTGTTCTAAACCTAGGGTTTACAAGAAGT 60
Db 1 tgttaagaagataaataattttgttttttcgctgttctaacaacctagggttacaagaagt 60

QY 61 AATTATCTGGAGCTAACAAATACTTTATTTTACCTTTTATTTTGCAGTAGTTTATGTT 120
Db 61 aatttctcgtgagctaaacaatactttatttttacctttttlatttgcagtagttttatgtt 120

QY 121 CAATTCATTAATTAATGATATTAATAAATTCCTCTGCAAAATATGAGGAGGACCTCATA 180
Db 121 caattctcaattcaatgataataaaatctcctcgcaaatatgtagagggagacctcata 180

QY 181 AAATATGTCATATGGAATGAGCAGATATAAAGATATATAGCTTTTCTTTGTCAAAAGG 240
Db 181 aaatatgtcatatggaatgagcagataataaagattatagctttctttgtcaaaagg 240

QY 241 AGACTCAATATCTTTACTTTCATGAGGACATTTGTCACAAATGTTTCCGCCATATTCAT 300
Db 241 agactcaaatattcttactcttctcatgaggacattgtgacaaatgttccccccagaatcat 300

QY 301 CCGGGGAACCACTCTGCGCCCATGATGCGCCCTGGACAAAGCTCCTTTCTGTAATATTGA 360
Db 301 ccggggaaaccacctctggcccatgtatgcccctggacaaagctccctttctgtaattga 360

QY 361 GTCATCATGAGTGAAGAACGGTGCATATTTGGTGTCAAAGTGTCACTGAACATAAGGCTGA 420
Db 361 gtcatacagtgaagaaacgctgcataattgtgtcacaagtgctcaactgaactaaaggctga 420

QY 421 CTTCCACACACACATAAGTGTGATAAATACTAAACAAAGAGATTTGGCATAAGTTGGTG 480
Db 421 cttccacagacacggaagtgtgataaaatctaaacaagaagaattggcataagttgggtg 480

QY 481 AATGTTTTTAAACATCCAAATTCATAGGCTTATAAATATTAATGCTATATTTTATCAA 540
Db 481 aatgttttttaaacatcccaatttcataaggcttataaataatttaattgctatattttatcaa 540

QY 541 CGAATCTCCAGTTGCTTTGCTGATGATGAGAAAGATATAAAGAAAGAAAGCTCAAGA 600
Db 541 cgaatctccagttgctttgtctgatgatagaaagataaaaaagaaagaaagctcaaga 600

QY 601 ACTCATATAAACCACACAAATGTAAGCTGTGTTATAAATGGGTGCCATGTAGATGGAA 660
Db 601 actcataaaacccacacaaatgtgaagctctgttataaatgggtgccaatgaagatggaa 660

QY 661 GAAGTATCTACATAGCAGAAGAGAGAAATGAATACTCATTTTATTCAGTTGGCCCC 720
Db 661 gaagtatctacataagcagaaggaagagaaatgaaatactcoatttttattgagttggcccc 720

QY 721 CACTGTATGTGGTGGTGTATTATTAAGGTGATGACCCAGGAGAAATTTGTAACATATAAA 780
Db 721 cactgtatgtggctgggtatttatgaagtgatgacccaggagaagaattgtaaacataaaa 780

QY 781 CCATCTCAATATATAACCCGAGGAGCAGACGATATCTCTATGAGCCCTGTATTTTACT 840
Db 781 ccactcacaataataaacccgaggcagaagcagcatatctctatgaagcctgtattact 840

QY 841 CAGTGGGAAATAAATTTATTAA 861
Db 841 cagtgggaaataaattttattaa 861

RESULT 2
AAAX22903
ID AAX22903 standard; DNA; 3951 BP.
XX AAX22903;
XX
DT 28-MAY-1999 (first entry)
XX Human DPD DNA.
DE
XX
KW DPD; dihydro:pyrimidine-dehydrogenase; monoclonal antibody; MAB;
KW immunoassay reagent; cancer patient; treatment; antitumour agent;
KW 5-fluorouracil; affinity purification; toxicity; ss.
OS Homo sapiens.
XX
PN DE19837391-A1.
XX
PD 25-FEB-1999.
XX
PF 18-AUG-1998; 98DE-1037391.
XX
PR 22-AUG-1997; 97EP-0114630.
XX
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX Hasegawa M, Yoshikubo T;
XX
XX WPI; 1999-155202/14.
DR P-PSDB; AAM93361.
XX
PT Monoclonal antibody specific for dihydro:pyrimidine dehydrogenase -
PS for assessing patient response to 5-fluorouracil antitumor agents
XX Disclosure; Page 15-17; 34pp; German.
XX
CC This invention describes a monoclonal antibody (MAB) specific for
CC dehydro:pyrimidine dehydrogenase (DPD). This MAB is used as immunoassay
CC reagents to identify a lack of DPD in a patient and to assess the
CC sensitivity of cancer patients to treatment with antitumor agents of the
CC 5-fluorouracil (5-FU) type. The MAB can also be used for affinity
CC purification of DPD. DPD is involved in reduction of 5-FU (and related
CC catabolites and derivatives) and lack of it is associated with increased
CC toxicity of this type of antitumor agent. It has specific binding
CC interaction. The MAB provide a sensitive and reliable test for DPD,
CC which is simple, rapid and suitable for routine screening.
XX
SQ Sequence 3951 BP; 1153 A; 785 C; 896 G; 1117 T; 0 other;

Query Match 20.1%; Score 172.8; DB 20; Length 3951;
Best Local Similarity 93.8%; Pred. No. 1.5e-29;
Matches 180; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 244 CTCATATCTTTTACTCTTTCATGAGGACATTTGTGACAAATGTTTCCCCCAATATCATCCG 303
Db 1797 caccaaacttctctcttgataagacattgtgacaaatgtttccccagaatcatccg 1856

QY 304 GGAACCACTCTGGCCCCATGTATGCCCTGGACAAAGCTCCTTTCTGTAATATTGACCT 363
Db 304 ggaacacactctggccccatgtatgccctggacaaagctcctttctgtaatatattgacct 363

CC This sequence represents the DNA sequence encoding human
CC dihydropyrimidine dehydrogenase (DPD). DPD catalyses the initial
CC and rate limiting step in pyrimidine catabolism, the reduction of
CC pyrimidines to 5,6-dihydropyrimidines. DPD is a complex enzyme
CC consisting of two identical subunits, containing FMN, FAD and iron-
CC sulphur centers, and utilising NADPH as a cofactor. DPD has also been
CC shown to catalyse the reduction of various pyrimidine analogues
CC including the fluoropyrimidine anticancer drug 5-fluorouracil (Fura).
CC Up to 85% of administered Fura may be catabolised by DPD, and it
CC therefore governs the effectiveness of Fura as an anticancer drug.
CC DPD genes or fragments of them may be used in the detection of DPD in a
CC sample, esp. isolated from a cancer patient. According to the amount
CC of DPD detected, a therapeutically effective amount of Fura may be
CC determined and administered. DPD deficiency, leading to life-
CC threatening toxicity on exposure to Fura, in a human caused by a
CC frameshift mutation may be determined by means of a molecular biological
CC assay to detect the deletion of an A residue at codon 318 within the
CC DPD-coding region.
XX

Sequence 4368 BP; 1325 A; 829 C; 932 G; 1282 T; 0 other;

Query Match 19.9%; Score 171.2; DB 16; Length 4368;
Best Local Similarity 93.2%; Pred. No. 3.5e-29;
Matches 179; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTCCCCCAATAATCATCCG 303
Db 1764 caccacaaattctctctgtataaggacattgtgacaaattttcccccagaatcatcgc 1823
QY 304 GGGAAACCACTCTGGCCCAATGATGGCCCTGGACAAAGTCCTTTCTGAAATATTGAGCT 363
Db 1824 gggaaaccaccttgcccccattgtatggccctggacaaagctcttctgaatttgagct 1883
QY 364 CATCAGTGAGAAACGGCTGCATATTGTGTCAAGTGTCACTGAACCTAAGGCTGACTT 423
Db 1884 catcagtgaagaaacggctgcattatgtgtgtaagtgctcaactgaactaaaggctgactt 1943
QY 424 CCCAGACAAACat 435
Db 1944 tccagacacat 1955

RESULT 5
AAT03143
XX ID AAT03143 standard; cDNA; 4369 BP.
XX AC AAT03143;
XX 05-JUN-1996 (first entry)
XX Human dihydropyrimidine dehydrogenase cDNA.
XX
XX Bovine: liver; human: dihydropyrimidine dehydrogenase; DPD;
KW pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;
KW fluoropyrimidine; anticancer drug; 5-fluorouracil; Fura; cancer;
KW frameshift mutation; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 49..3125
FT CDS
FT /*tag= a
FT /product= Human lymphocyte DPD
FT complement (36..60)
FT primer_bind
FT /*tag= b
FT complement (716..743)
FT primer_bind
FT /*tag= c
FT 1260..1288
FT /*tag= d
FT 3117..3139
FT primer_bind
FT /*tag= e
FT 1000
FT mutation

FT /*tag= f
FT /note= "Position of single base deletion in DPD
XX deficient patients"
XX
XX W09528489-A1.
XX
XX 26-OCT-1995.
XX
XX 13-APR-1995; 95WO-US04567.
XX
XX 13-APR-1994; 94US-0227357.
XX (UABR-) UAB RES FOUND.
XX Cheng X, Diasio RB, Johnson M, Lu Z, Zhang R;
XX WPI; 1995-373803/48.
XX P-PSDB; AAR86001.
XX
XX Novel dihydro:pyrimidine dehydrogenase gene - used to optimise
XX 5-fluoro:uracil doses given to cancer patients
XX
XX Disclosure; Fig 7; 207pp; English.
XX
XX This sequence represents the cDNA sequence encoding human dihydro-
XX pyrimidine dehydrogenase (DPD). This sequence was amplified using
XX the primer sequences given in AAT03144-47. DPD catalyses the initial
XX and rate limiting step in pyrimidine catabolism, the reduction of
XX pyrimidines to 5,6-dihydropyrimidines. DPD is a complex enzyme
XX consisting of two identical subunits, containing FMN, FAD and iron-
XX sulphur centers, and utilising NADPH as a cofactor. DPD has also been
XX shown to catalyse the reduction of various pyrimidine analogues
XX including the fluoropyrimidine anticancer drug 5-fluorouracil (Fura).
XX Up to 85% of administered Fura may be catabolised by DPD, and it
XX therefore governs the effectiveness of Fura as an anticancer drug.
XX DPD genes or fragments of them may be used in the detection of DPD in a
XX sample, esp. isolated from a cancer patient. According to the amount
XX of DPD detected, a therapeutically effective amount of Fura may be
XX determined and administered. DPD deficiency, leading to life-
XX threatening toxicity on exposure to Fura, in a human caused by a
XX frameshift mutation may be determined by means of a molecular biological
XX assay to detect the deletion of an A residue at codon 318 within the
XX DPD-coding region.
XX
XX Sequence 4369 BP; 1325 A; 829 C; 932 G; 1283 T; 0 other;

Query Match 19.9%; Score 171.2; DB 16; Length 4369;
Best Local Similarity 93.2%; Pred. No. 3.5e-29;
Matches 179; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 244 CTCAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTCCCCCAATAATCATCCG 303
Db 1764 caccacaaattctctctgtataaggacattgtgacaaattttcccccagaatcatcgc 1823
QY 304 GGGAAACCACTCTGGCCCAATGATGGCCCTGGACAAAGTCCTTTCTGAAATATTGAGCT 363
Db 1824 gggaaaccaccttgcccccattgtatggccctggacaaagctcttctgaatttgagct 1883
QY 364 CATCAGTGAGAAACGGCTGCATATTGTGTCAAGTGTCACTGAACCTAAGGCTGACTT 423
Db 1884 catcagtgaagaaacggctgcattatgtgtgtaagtgctcaactgaactaaaggctgactt 1943
QY 424 CCCAGACAAACat 435
Db 1944 tccagacacat 1955

RESULT 6
AAT03132
ID AAT03132 standard; DNA; 4414 BP.
XX
XX AAT03132;
AC

XX	04-JUN-1996	(first entry)
XX	Bovine dihydropyrimidine dehydrogenase gene.	
DE	Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;	
DE	pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;	
KW	fluoropyrimidine; anticancer drug; 5-fluorouracil; Fura; cancer;	
KW	frameshift mutation; ss.	
XX	Bos taurus.	
OS		
XX	Key Location/Qualifiers	
FH	68..3145	
FT	/tag= a	
FT	/product= DPD	
WT		
WT		
PN	WO9528489-A1.	
PD	26-OCT-1995.	
XX		
XX	13-APR-1995; 95WO-US04567.	
PF		
XX	13-APR-1994; 94US-0227357.	
PR		
XX	(UABR-) UAB RES FOUND.	
PA		
XX	Cheng X, Diasio RB, Johnson M, Lu Z, Zhang R;	
PI		
PI	WPI; 1995-373803/48.	
DR	P-PSDB; AAR86000.	
DR		
XX	Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise	
PT	5-fluoro:uracil doses given to cancer patients	
PT		
XX	Claim 4; Page 115-32; 207pp; English.	
PS		
XX	This sequence represents the DNA sequence encoding bovine liver	
CC	dihydropyrimidine dehydrogenase (DPD). DPD catalyses the initial	
CC	and rate limiting step in pyrimidine catabolism, the reduction of	
CC	pyrimidines to 5,6-dihydropyrimidines. DPD is a complex enzyme	
CC	consisting of two identical subunits, containing FMN, FAD and iron-	
CC	sulphur centers, and utilising NADPH as a cofactor. DPD has also been	
CC	shown to catalyse the reduction of various pyrimidine analogues	
CC	including the fluoropyrimidine anticancer drug 5-fluorouracil (Fura).	
CC	Up to 85% of administered Fura may be catabolised by DPD, and it	
CC	therefore governs the effectiveness of Fura as an anticancer drug.	
CC	DPD genes or fragments of them may be used in the detection of DPD in a	
CC	sample, esp. isolated from a cancer patient. According to the amount	
CC	of DPD detected, a therapeutically effective amount of Fura may be	
CC	determined and administered. DPD deficiency, leading to life-	
CC	threatening toxicity on exposure to Fura, in a human caused by a	
CC	frameshift mutation may be determined by means of a molecular biological	
CC	assay to detect the deletion of an A residue at codon 318 within the	
CC	DPD-coding region.	
XX		
SQ	Sequence 4414 BP; 1315 A; 872 C; 956 G; 1271 T; 0 other;	
	Query Match 18.0%; Score 154.8; DB 16; Length 4414;	
	Best Local Similarity 88.4%; Pred. No. 1.5e-25;	
	Matches 168; Conservative 0; Mismatches 22; Indels 0; Gaps 0;	
QY	246 CAATATCTTTACTTTCATGAGACATTGTGCACAATCTTCCCCCAATAATCATCGGG 305	
Db	1785 ccacaaacttctctctaggaataggtgacaattttccaccagaatacctcggg 1844	
QY	306 GAACACCCTCTGCCCCCATGTATFGGCCCTGGACAAAGCTCTTCTGAATATGAGCTCA 365	
Db	1845 ggaccaccttgcccccatgatggcccctggacaaaagtcttctctgaattaggactca 1904	
QY	366 TCAGTGAGAAAACGGGTGCATATTTGTGTCAAGTGTCACTGAACTAATAGGCTGCTCC 425	

Db	1905	tcagtgaataaacgcgtgcataattggtgtccaagtgtcactgaactaaagccgacttc	1964
QY	426	CAGACACAT 435	
Db	1965	cagacaatat	1974
RESULT	7		
AAT03142			
ID	AAT03142	standard; DNA; 4430 BP.	
XX			
AC	AAT03142;		
XX			
DT	04-JUN-1996	(first entry)	
XX			
DE	Bovine dihydropyrimidine dehydrogenase cDNA.		
XX			
KW	Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;		
KW	pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;		
KW	fluoropyrimidine; anticancer drug; 5-fluorouracil; Fura; cancer;		
XX	frameshift mutation; ss.		
OS	Bos taurus.		
XX			
FH	Key Location/Qualifiers		
FT	67..3144		
FT	/tag= a		
FT	/product= Bovine DPD		
FT	complement (2032..2051)		
FT	/tag= b		
FT	/note= "Primer A"		
FT	2080..2096		
FT	/tag= c		
FT	/note= "Primer B"		
FT	complement (2035..2078)		
FT	/tag= d		
FT	/note= "Primer C"		
FT	complement (85..104)		
FT	/tag= e		
FT	/note= "Primer D"		
FT	2141..2160		
FT	/tag= f		
FT	/note= "Primer E"		
FT	210..229		
FT	/tag= g		
FT	/note= "Primer F"		
XX	WO9528489-A1.		
XX	26-OCT-1995.		
XX	13-APR-1995; 95WO-US04567.		
PR	13-APR-1994; 94US-0227357.		
XX	(UABR-) UAB RES FOUND.		
PI	Cheng X, Diasio RB, Johnson M, Lu Z, Zhang R;		
XX	WPI; 1995-373803/48.		
DR	P-PSDB; AAR86000.		
XX	Novel di:hydro:pyrimidine dehydrogenase gene - used to optimise		
PT	5-fluoro:uracil doses given to cancer patients		
PT			
XX	Claim 4; Page 115-32; 207pp; English.		
XX	This sequence represents the DNA sequence encoding bovine liver		
CC	dihydropyrimidine dehydrogenase (DPD). DPD catalyses the initial		
CC	and rate limiting step in pyrimidine catabolism, the reduction of		
CC	pyrimidines to 5,6-dihydropyrimidines. DPD is a complex enzyme		
CC	consisting of two identical subunits, containing FMN, FAD and iron-		
CC	sulphur centers, and utilising NADPH as a cofactor. DPD has also been		
CC	shown to catalyse the reduction of various pyrimidine analogues		
CC	including the fluoropyrimidine anticancer drug 5-fluorouracil (Fura).		
CC	Up to 85% of administered Fura may be catabolised by DPD, and it		
CC	therefore governs the effectiveness of Fura as an anticancer drug.		
CC	DPD genes or fragments of them may be used in the detection of DPD in a		
CC	sample, esp. isolated from a cancer patient. According to the amount		
CC	of DPD detected, a therapeutically effective amount of Fura may be		
CC	determined and administered. DPD deficiency, leading to life-		

	Query Match	18.0%	Score 154.8;	DB 16;	Length 4414;
	Best Local Similarity	88.4%;	Pred. No. 1.5e-25;		
	Matches 168;	Conservative 0;	Mismatches 22;	Indels 0;	Gaps 0;
QY	246	CAATATCTTACTCTTTCATGAGGACATATGACAAATGTTTCCCCCATATAATCATCCGGG	305		
Db	1785	ccaaacttctctctgataagatagtgcaaatgtttccaccagaatcatccggg	1844		
QY	306	GAACCACTCTGGCCCCCATGTATGGCCCTGCACAAAGCTCCTTTCTGTAATATTGAGCTCA	365		
Db	1845	ggaccacctctgcccacatgtatggccctggacaaaagtctttcctgaattagactca	1904		
QY	366	TCAGTGAGAAACGGCTGCATATTTGGTGTCAAAGTGTCACTGAACCTAAAGCGTGACTTCC	425		


```
AAF58254
ID  AAF58254 standard; DNA; 936 BP.
XX
AC  AAF58254;
XX
DT  24-APR-2001 (first entry)
XX
DE  Oligonucleotide D1875.
XX
KW  Electron-transfer group; ETM; mismatch; genotyping;
KW  gene expression; ss.
XX
OS  Synthetic.
XX
PN  WO200107665-A2.
XX
PD  01-FEB-2001.
XX
PF  26-JUL-2000; 2000WO-US20476.
XX
PR  26-JUL-1999; 99US-0145695.
PR  17-MAR-2000; 2000US-0190259.
XX
PA  (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI  Umek RM;
XX
DR  WPI; 2001-159728/16.
XX
PT  Nucleic acids containing electron-transfer group, useful as labels in
PT  hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT  a single surface
XX
PS  Example 6; Page 127; 159pp; English.
XX
CC  The present invention relates to a composition comprising two nucleic
CC  acids each containing an electron-transfer group (ETM) having
CC  different redox potentials. The invention is used for electronic
CC  detection of nucleic acids, especially of substitutions (mismatches)
CC  and single-nucleotide polymorphisms, e.g. for genotyping,
CC  monitoring gene expression.
XX
SQ  Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match      17.7%; Score 152.6; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 3.4e-25;
Matches 8; Conservative 512; Mismatches 271; Indels 0; Gaps 0;

QY  9 AAGATAAATATTTTGGTTTCGCTGTTCTTAAACCTAGGGTTACAAGAAGTAATTTATC 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  2 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61

QY  69 TGGAGCTAACAATACTTTATTTACCTTTTATTTGCAAGTAGTTTATGTTCAATTTCTA 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  62 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 121

QY  129 ATTTAATGCTATATTAATAATTCCTCTGCAATATGTGAGGAGGACCTCATATAAATATTG 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  122 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181

QY  189 TCATATGGAATGAGCAGATAATAAGATTATAGCTTTTCTTCTCAAAAGGAGACTCA 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  182 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241

QY  249 TATCTTTACTCTTCATGAGGACATTTGCAAAATGTTTCCCCCATATATCATCCGGGAA 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  242 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 301

QY  309 CCACCTCTGGCCCCCATGTATGCGCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCATCA 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  302 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361
```

RESULT 10

AAF58257

ID AAF58257 standard; DNA; 936 BP.

XX

AC AAF58257;

XX

DT 24-APR-2001 (first entry)

XX

DE Oligonucleotide D1954.

XX

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX

OS Synthetic.

XX

PN WO200107665-A2.

XX

PD 01-FEB-2001.

XX

PF 26-JUL-2000; 2000WO-US20476.

XX

PR 26-JUL-1999; 99US-0145695.

PR

17-MAR-2000; 2000US-0190259.

XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX

PI Umek RM;

XX

DR WPI; 2001-159728/16.

XX

PT Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

PT a single surface

XX

PS Example 6; Page 127; 159pp; English.

XX

CC The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.

SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 17.7%; Score 152.6; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 3.4e-25;
Matches 8; Conservative 512; Mismatches 271; Indels 0; Gaps 0;

```
QY 9 AAGATAAATATTTTGTCTTTCGCTGTCTTAACCTAGGTTACAGAAGTAATTTATC 68
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  61
D  2  wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 61
QY 69 TGGAGCTAACAAATACTTTATTTTACCTTTTATTGCAAGTACTTTATGTTCAATCTA 128
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  121
D  62  wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 121
QY 129 ATTTAATGTATTTAAATTCCTCTGCAAAATATGTGAGGAGGACCTCATAAATATTG 188
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  181
D  122 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 181
QY 189 TCATATGGAATGAGCAGATAATAAGATTATAGCTTTTCTTGCAAAAGGAGACTCAA 248
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  241
D  182 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 241
QY 249 TATCTTTTACTCTTTCATGAGGACATTTGTGACAAATGTTTCCCCCATATATATCCGGGAA 308
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  301
D  242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 301
QY 309 CCACCTCTGGCCCCCATGTATGGCCCTGGACAAAGCTCCTTCTGAATATTGAGCTCATCA 368
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  361
D  302 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 361
QY 369 GTGAGAAAGCGGTGCATATTGTTGTCAAAGTGTCACTGAAGTCAAAAGGCTGACTCCAG 428
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  421
D  362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 421
QY 429 ACAACaTAAGTGTGATAAAATCTAAAACAAGAAATGGCATAGTTGTGAATGTTTA 488
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  481
D  422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 481
QY 489 TTTAAACATCCAAATTCATAGGCTTATAAATATTAATGTGTATATTTTATCAACGAATCTG 548
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  541
D  482 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 541
QY 549 CCAGTGTCTTCTGATGCATAGAAAGATAAAAGAAAGAAAGCTCAAGAACTCAAA 608
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  601
D  542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 601
QY 609 AAACCCACACATGTGAAGCTCTGTATATAAATGGTGCCATGTGAAGTGAAGAAGTATC 668
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  661
D  602 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 661
QY 669 TACATAAGCAGAGAGAGAAATGAATACTCATTTTATTTGAGTTGGCCCCCACTGAT 728
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  721
D  662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 721
QY 729 GTGGCTGTTATTATGAGGTGATGACCCAGGAAGAAATTTGTAACATATAAACCACTCCA 788
D  :::::::::::::::::::::  ::::::::::  ::::::::::  ::::::::::  ::::::::::  781
D  722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 781
QY 789 AATATAAACCC 799
D  ::::::::::  |||
D  782 wwwwwwgcccc 792
```

RESULT 11

AAF58259

ID AAF58259 standard; DNA; 936 BP.

XX

AC AAF58259;

XX

DT 24-APR-2001 (first entry)

XX Oligonucleotide D2004.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

OS WO200107665-A2.

PN 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

PA Umek RM;

XX WPI; 2001-159728/16.

DR Nucleic acids containing electron-transfer group, useful as labels in

XX hybridization assays, e.g. for genotyping, allowing repeat analyses on

PT a single surface

XX Example 6; Page 128; 159pp; English.

PS The present invention relates to a composition comprising two nucleic

XX acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,

CC monitoring gene expression.

XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 17.7%; Score 152.6; DB 22; Length 936;

Best Local Similarity 1.0%; Pred. No. 3.4e-25;

Matches 8; Conservative 512; Mismatches 271; Indels 0; Gaps 0;

QY 9 AAGATAAATATTTTGTCTTTCGCTGTCTTAACCTAGGTTACAGAAGTAATTTATC 68

D ::::::::::::::::::::: :::::::::: :::::::::: :::::::::: :::::::::: 61

D 2 ww 61

QY 69 TGGAGCTAACAAATACTTTTATTTTACCTTTTATTGCAAGTACTTTTATGTTCAATCTA 128

D ::::::::::::::::::::: :::::::::: :::::::::: :::::::::: :::::::::: 121

D 62 ww 121

QY 129 ATTTAATGTATTTAAATTCCTCTGCAAAATATGTGAGGAGGACCTCATAAATATTG 188

D ::::::::::::::::::::: :::::::::: :::::::::: :::::::::: :::::::::: 181

D 122 ww 181

QY 189 TCATATGGAATGAGCAGATAATAAGATTATAGCTTTTCTTGCAAAAGGAGACTCAA 248

D ::::::::::::::::::::: :::::::::: :::::::::: :::::::::: :::::::::: 241

D 182 ww 241

QY 249 TATCTTTTACTCTTTCATGAGGACATTTGTGACAAATGTTTCCCCCATATATATCCGGGAA 308

D ::::::::::::::::::::: :::::::::: :::::::::: :::::::::: :::::::::: 301

D 242 ww 301

QY 309 CCACCTCTGGCCCCCATGTATGGCCCTGGACAAAGCTCCTTCTGAATATTGAGCTCATCA 368

D ::::::::::::::::::::: :::::::::: :::::::::: :::::::::: :::::::::: 361

D 302 ww 361

QY 369 GTGAGAAAGCGGTGCATATTGTTGTCAAAGTGTCACTGAAGTCAAAAGGCTGACTCCAG 428

D ::::::::::::::::::::: :::::::::: :::::::::: :::::::::: :::::::::: 421

D 362 ww 421

QY 429 ACAACaTAAGTGTGATAAAATCTAAAACAAGAAATGGCATAGTTGGTGAATGTTTA 488

CC 4-fluorouracil.

Sequence 4447 BP; 1285 A; 928 C; 1000 G; 1234 T; 0 other;
SQ
XX

Query Match	17.6%	Score 151.6;	DB 17;	Length 4447;
Best Local Similarity	87.4%;	Pred. No. 7.9e-25;		
Matches 166;	Conservative	0;	Mismatches 25;	Indels 0;
			Gaps	0;

Qy	246	CAATATCTTTACTCTTTTCATGAGGACATTCTGACAAATGTTTCCCCATAATCATCGGG	305
Db	1805	ccaaaactttctctcttgataaggacatactgacaaaatgtctaccagaatcgtccggg	1864

Qy 306 GAACCACTCTGGCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCTCA 365

uu 1005 ggacacacccctggccccctgacggccctggacaaagctcctctctgaataattgagatca 1922

Qy 366 TCAGTGAGAAAAACGGCTGCATATTGGTGTCAAAGTGTCACTGAACATAAAGGCTGACTTCC 425

Db	1925	tcagtgaaaaacagctgcataatgggtcgaaggtcactgaactaaaagctgacttc	1984
Qy	426	CAGACAAcAT	435

Db 1985 cagacaatat 1994

RESULT 15
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.

XX
AC
AAF58252;
XX
DT
2A-APP-2001 /fig+ extra)

DI 24-APR-2001 (11P5C EMU1)
XX
XX
DE Oligonucleotide D1835.
XX

KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
XX
XX

OS	Synthetic.
XX	
PN	WO200107665-A2.
XX	

PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.

XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
YY

AA
PA
XX
PI (CLIN-) CLINICAL MICRO SENSORS INC.
PI Umek RM;

XX
DR WPI; 2001-159728/16.
XX
XX

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX

Example 6; Page 127; 159pp; English.
The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms (SNPs) for sequencing
CC

CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 17.3%; Score 148.6; DB 22; Length 936;

[illegible]

Search completed: November 20, 2001, 16:49:18
Job time: 4632 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 15:07:46 ; Search time 1366.05 Seconds
(without alignments)
5957.986 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTATGAAGATAAATATT.....ACTGGGAATAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

- 1: gb_estl1.*
- 2: gb_estl2.*
- 3: gb_estl3.*
- 4: gb_estl4.*
- 5: gb_estl5.*
- 6: gb_estl6.*
- 7: gb_estl7.*
- 8: gb_estl8.*
- 9: gb_estl9.*
- 10: gb_estl10.*
- 11: gb_estl11.*
- 12: gb_estl12.*
- 13: gb_estl13.*
- 14: gb_estl14.*
- 15: gb_estl15.*
- 16: gb_estl16.*
- 17: gb_estl17.*
- 18: gb_estl18.*
- 19: gb_estl19.*
- 20: gb_estl20.*
- 21: gb_estl21.*
- 22: gb_estl22.*
- 23: gb_estl23.*
- 24: gb_estl24.*
- 25: gb_estl25.*
- 26: gb_estl26.*
- 27: gb_estl27.*
- 28: gb_estl28.*
- 29: gb_estl29.*
- 30: gb_estl30.*
- 31: gb_estl31.*
- 32: gb_estl32.*
- 33: gb_estl33.*
- 34: gb_estl34.*
- 35: gb_estl35.*
- 36: gb_estl36.*
- 37: gb_estl37.*
- 38: gb_estl38.*
- 39: gb_estl39.*
- 40: gb_estl40.*
- 41: gb_estl41.*
- 42: gb_estl42.*
- 43: gb_estl43.*
- 44: gb_estl44.*
- 45: gb_estl45.*
- 46: gb_estl46.*
- 47: gb_estl47.*

- 44: em_esthum10.*
- 45: em_esthum11.*
- 46: em_esthum12.*
- 47: em_esthum13.*
- 48: em_esthum14.*
- 49: em_esthum15.*
- 50: em_esthum16.*
- 51: em_esthum17.*
- 52: em_esthum18.*
- 53: em_esthum19.*
- 54: em_esthum20.*
- 55: em_esthum21.*
- 56: em_esthum22.*
- 57: em_esthum23.*
- 58: em_esthum24.*
- 59: em_esthum25.*
- 60: em_esthum26.*
- 61: em_esthum27.*
- 62: em_esthum28.*
- 63: em_estin1.*
- 64: em_estin2.*
- 65: em_estin3.*
- 66: em_estin4.*
- 67: em_estin5.*
- 68: em_estom1.*
- 69: em_estom2.*
- 70: em_estov1.*
- 71: em_estov2.*
- 72: em_estpl1.*
- 73: em_estpl2.*
- 74: em_estpl3.*
- 75: em_estpl4.*
- 76: em_estpl5.*
- 77: em_estpl6.*
- 78: em_estpl7.*
- 79: em_estpl8.*
- 80: em_estpl9.*
- 81: em_estpl10.*
- 82: em_estro1.*
- 83: em_estro2.*
- 84: em_estro3.*
- 85: em_estro4.*
- 86: em_estro5.*
- 87: em_estro6.*
- 88: em_estro7.*
- 89: em_estro8.*
- 90: em_estro9.*
- 91: em_estro10.*
- 92: em_estro11.*
- 93: em_estro12.*
- 94: em_estro13.*
- 95: em_estro14.*
- 96: em_estro15.*
- 97: em_estro16.*
- 98: em_estro17.*
- 99: em_estro18.*
- 100: em_estro19.*
- 101: em_estro20.*
- 102: gb_est25.*
- 103: gb_est26.*
- 104: gb_est27.*
- 105: gb_est28.*
- 106: gb_est29.*
- 107: gb_est30.*
- 108: gb_est31.*
- 109: gb_est32.*
- 110: gb_est41.*
- 111: gb_est42.*
- 112: gb_est43.*
- 113: gb_est44.*
- 114: gb_est45.*
- 115: gb_est46.*
- 116: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

TITLE Saurin, W. and Weissenbach, J.
JOURNAL Human gene number estimate provided by genome wide analysis using
REFERENCE Tetraodon nigroviridis DNA sequence
AUTHORS 3 (bases 1 to 853)
TITLE Genoscope.
JOURNAL Direct Submission
COMMENT Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.

FEATURES

source
 Location/Qualifiers
 1..853
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="262A08"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG262BA04LP1-end : T7"

BASE COUNT 252 a 189 c 136 g 269 t 7 others
 ORIGIN

Query Match 10.58; Score 90.8; DB 220; Length 853;
 Best Local Similarity 67.9%; Pred. No. 1e-10;
 Matches 125; Conservative 1; Mismatches 58; Indels 0; Gaps 0;

QY 258 TCTTTTCATGAGGACATCTTGACAAATGTTTCCCCCAATATCATCCGGGGAACCACTCTG 317

Db 409 TGTTTCTCAGAGCTTGTAACAATGTCTCTCGTATGTGCGAGGACCACATTCG 350

QY 318 GCCCATGATGGCCCTGGACAAAGCTCTTTCTGAATATTGAGCTCATCAGTGAGAAAA 377

Db 349 GCAATGTGTATGGCCAGGTGAGGGTCATTCTCTCAACATTGAGCTTATTAGCGAGA 290

QY 378 CGGCTGCATATTGGTGCAAGTGTCTACTGAACATAAGGCTGATTTCCACAGAACATAA 437

Db 289 CGCGAGCTTATTGGTGTCAATCAGTGGCGMACTTAAGAGGAGCTTTCCGAATATGTA 230

QY 438 GTGT 441

Db 229 GTTT 226

RESULT 3

AW018806
 LOCUS AW018806 621 bp mRNA EST 13-SEP-1999
 DEFINITION fd60d03.y1 zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
 SW:DPYD_HUMAN Q12882 DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+]

PRECURSOR i; mRNA sequence.

AW018806

AW018806.1 GI:5872336

EST.

zebrafish.

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Rasbora; Danio.

1 (bases 1 to 621)

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy

, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood

, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,

Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.

and Willson, R.

WashU Zebrafish EST Project 1998

Unpublished (1998)

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrfishnewatson.wustl.edu

cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 RessourcenZentrumPrimaridenbank, Berlin, Germany (web address:
 www.rzpd.de)

Seq primer: T3 ET from Amersham

High quality sequence stop: 491.

FEATURES

source
 Location/Qualifiers
 1..621
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish WashU MPIMG EST"
 /sex="mixed"
 /tissue_type="26 somite embryos, adult livers, shield
 stage embryos"
 /lab_host="XLI-blue MRP"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
 strand cDNA was primed with a Not I - oligo(dT)15 primer
 15'pGACTAGTTCTAGATCGGAGCGCGCCCTTTTCTTTT3';
 double-stranded cDNA was ligated to Sal I adaptors (BRL),
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late mitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

BASE COUNT 151 a 169 c 160 g 141 t

ORIGIN

Query Match 10.2%; Score 87.8; DB 110; Length 621;

Best Local Similarity 71.2%; Pred. No. 4.9e-10;

Matches 116; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 275 GTGACAAATGTTTCCCCCAATATCATCCGGGGAACCACTCTGCGCCCATGTATGGCCT 334

Db 1 GTGACCAACGTGTCAACCGCAATCGTGGGACACCTTCAGGTACATATTCGGTCT 60

QY 335 GGACAAAGCTCCTTTCTGAATATTGAGCTCATCAGTGAGAAAGCGGTGATATTTGTT 394

Db 61 GGTGAGGGCTCATCTCTCAACATCGAGCTCATTTAGTAAAAAAGCTGCGCTTACTGTGC 120

QY 395 CAAAGTGTCACTGAACATAAGGCTGACTTCCACAGACACATAA 437

Db 121 AGAGCGTGCTGAGTCAAAAGCCGACTTCCCAAAAATATTA 163

RESULT 4

CNS00175

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR03G12 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL058798

AL058798.1 GI:4939103

GSS.

KEYWORDS

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1207)

CNS00175

1207 bp

DNA

GSS

04-JUN-1999

AUTHORS
TITLE Genoscope.
JOURNAL Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1. 1207
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACRO3G12"
 /notes="end : TET3"

BASE COUNT 250 a 298 c 328 g 233 t 98 others
ORIGIN

Query Match 8.3%; Score 71.4; DB 219; Length 1207;
 Best Local Similarity 63.9%; Pred. No. 2.8e-06;
 Matches 108; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 267 AGGACATTGTGACAAATGTTTCCCGGCAATATCATCCGGGAACACCTCTGGCCCCCATGT 326
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 399 AGGATCTGGTCACGAATGTCTCGCGGCATCTGTCAGGGGCACCACTCGGGCTACAAAT 458
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 327 ATGCGCCCTGGCAAGAGCTCTTCTGAATATTGAGCTCATCAGTGAGAAACGGCTCAT 386
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 459 ATGACCCGACGACGGGCTGTTCTGAACATCGAATCTCGGGAAGGGGCCGAGT 518
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 387 ATTGGTCTCAAGGTGTCACCTGAACATAAGGCTGACTTCCCGACACAAT 435
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 519 ACTGGCTGAATCGATTGGAGAACTGAAGCGTGACTTCCCGGAGAAGAT 567
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 5
BE288474
LOCUS BE288474 562 bp mRNA EST 26-OCT-2000
DEFINITION 601095732F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3490423 5',
 mRNA sequence.
ACCESSION BE288474
VERSION BE288474.1 GI:9168194
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 562)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

FEATURES
 source
 1. 562
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone="IMAGE:3490423"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /notes="Organ: mammary; Vector: pCMV-SPORT6; Site.1: Sall;
 Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 137 a 124 c 166 g 135 t
ORIGIN

Query Match 7.4%; Score 64; DB 165; Length 562;
 Best Local Similarity 93.1%; Pred. No. 0.00013;
 Matches 67; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 364 CATCAGTGAGAAAACGGCTGCATATTGTTGTTCAAGTGTCTCACTGAACATAAGGCTGACTT 423
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1 CATCAGTGAGAAAACAGCTGCATATTGTTGTTCAAGTGTCTCACTGAACATAAGGCTGACTT 60
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 424 CCCAGACAACat 435
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 61 CCCGGACAATAT 72
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 6
CNS0039G
LOCUS CNS0039G 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACRO8K10 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 AL063921
 AL063921.1 GI:4941778
 GSS.
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 Genoscope.
 1 (bases 1 to 1101)
REFERENCE Direct Submission
AUTHORS BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

BACKWARD: GTTTCCTCCAGTCCAGCAGC
Plate: 70 row: B column: 4
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
Source
1..483
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 128 a 105 c 138 g 112 t

Query Match 6.4%; Score 54.8; DB 151; Length 483;
Best Local Similarity 89.7%; Pred. No. 0.017;
Matches 70; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
Qy 359 GAGTCTCATCTG-AGAAACGGTGCATATTGGTCAAAAGTGTCACTGAACATAAGGC 417
Db 1 CGGCTCATCTGAGAAACGGTGCATATTGGTCAAAAGTGTCACTGAACATAAGGC 60
Qy 418 TGACTTCCACAGCAAT 435
Db 61 CGACTTCCACAAATAT 78

RESULT 9
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/note="BACR29B23"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN
Query Match 6.1%; Score 52.6; DB 219; Length 1101;
Best Local Similarity 35.7%; Pred. No. 0.055;
Matches 99; Conservative 51; Mismatches 127; Indels 0; Gaps 0;
Qy 1 TGTAAATGACATAAATATTTTTCGCTGCTCTAAACCTAGGTTACAGAGT 60
Db 723 TATATATATTTTAAATTTTTCGCTGCTCTAAACCTAGGTTACAGAGT 664
Qy 61 AATTATCTGAGCTAACAAATACCTTTTATTTTACCTTTTATTTGCAAGTAGTTATGTT 120
Db 663 WATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 604
Qy 121 CAATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 180
Db 603 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 544
Qy 181 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
Db 543 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 484
Qy 241 AGACTCAATATCTTTTACTCTTTTCTCATGAGGACATTTGTG 277
Db 483 TWTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 447

RESULT 10
AZ527678
LOCUS AZ527678 912 bp DNA GSS 03-NOV-2000
DEFINITION ENTBN23TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION AZ527678
VERSION AZ527678.1 GI:11079801
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 912)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 20
High quality sequence stop: 734.
Location/Qualifiers
1..912
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,

FEATURES
source
1..912
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,

/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 264 a 123 c 59 g 435 t
ORIGIN

Query Match 6.0%; Score 51.8; DB 247; Length 881;
Best Local Similarity 50.2%; Pred. No. 0.083;
Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 444 TAAATCTTAAACAGAGAAATGGCATAAGTTGGTGAATGTTTATTTAAACATCCCAATT 503
Db 619 TTAATATATTAACATAACAAACATTCAGAAATAAGAAATGAATATTTAAACAATGAAAGT 560
QY 504 CATAGGCTTATAAATTAATGTGTATATTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
Db 559 GTTGATTTATTTAATATTTGTTCAACATAAATTCATTAACATATATTTTGATAATAT 500
QY 564 ATGCATAGAAAGATAAAGAAAGAAAGTCAAGAACTCATAAACCCACACAATGT 623
Db 499 ATAAATGGAATATTAAGGAATATACAATCTCAAGAACTAATAAAAGAAATAAAAG 440
QY 624 GAAGCTCTGTATTAATGGTCCCTGTGAAGTGAAGAACTCTACATAAGCAGAGG 683
Db 439 AATGAATAAATGAAGATAAATTTGTAATAATAACAGAAAGAAATATAAACGACAAAGA 380
QY 684 AAGAGAAATGAAATA 698
Db 379 AAAGAATAATAGA 365

RESULT 15
A2542627
LOCUS ENTXX21TR Entamoeba histolytica Sheared DNA Entamoeba histolytica 14-NOV-2000
DEFINITION genomic, DNA sequence.
ACCESSION A2542627
VERSION A2542627.1 GI:11160786
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 899)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
JOURNAL
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 93
High quality sequence stop: 858.
Location/Qualifiers

source

1. .899
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="vector: pHOSt; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 429 a 61 c 150 g 259 t
ORIGIN

Query Match 6.0%; Score 51.8; DB 245; Length 899;
Best Local Similarity 50.2%; Pred. No. 0.083;
Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 444 TAAAAATCTAAACAGAGAAATGGCATAAGTTGGTGAATGTTTATTTAAACATCCCAATT 503
Db 371 TTAATATATTAACATAACAAACATTTTCAGAAATAAGAAATGAATATTTAAACAATGAAAGT 430
QY 504 CATAGGCTTATAAATTAATGTGTATATTTTATCAACGAATCTGCCAGTTGCTTTGCTG 563
Db 431 GTTGATTTATTTAATATTTGTTCAACATAAATTCATAAACAATATATTTTGATAATAT 490
QY 564 ATGCATAGAAAGATAAAGAAAGAAAGCTCAAGAACTCATAAACCCACACAATGT 623
Db 491 ATAAATGGAATATTTAAAGGAATATACAACTCAAGAACTAATAAAAGAAATAAAAG 550
QY 624 GAAGCTCTGTATTAATGAATGGTGGCCATGTAAAGATGAAGAACTATCTACATAAGCAGAGG 683
Db 551 AATGAATAAATGAAGATAAATTTTGATAATAATACAGAAAGAAATATAAACGACAAAGA 610
QY 684 AAGAGAAATGAAATA 698
Db 611 AAAGAATAATAGA 625

Search completed: November 20, 2001, 16:21:54
Job time: 4448 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 14:27:41 ; Search time 1472.19 Seconds
(without alignments)
9046.205 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TCTTAATGAAGATAAATATT.....ACTGGGAATAAATTTATTAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba1:*

2: gb_ba2:*

3: gb_ba3:*

4: gb_in1:*

5: gb_in2:*

6: gb_in3:*

7: gb_om:*

8: gb_ov:*

9: gb_pat1:*

10: gb_pat2:*

11: gb_ph:*

12: gb_pl1:*

13: gb_pl2:*

14: gb_pl3:*

15: gb_pl4:*

16: em_ba1:*

17: em_ba2:*

18: em_fun:*

19: em_htgo_hum:*

20: em_htgo_inv:*

21: em_htgo_rod:*

22: em_htg_hum1:*

23: em_htg_hum2:*

24: em_htg_hum3:*

25: em_htg_hum4:*

26: em_htg_hum5:*

27: em_htg_hum6:*

28: em_htg_hum7:*

29: em_htg_hum8:*

30: em_htg_inv1:*

31: em_htg_inv2:*

32: em_htg_other:*

33: em_htg_rod:*

34: em_hum1:*

35: em_hum2:*

36: em_hum3:*

37: em_hum4:*

38: em_hum5:*

39: em_hum6:*

40: em_hum7:*

41: em_in:*

42: em_om:*

43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_vil2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pri:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_rol:*

95: gb_rol2:*

96: gb_in4:*

97: gb_pri0:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	775.2	90.0	180286	80	AL356574 Homo sapi
2	608.6	70.7	626	97	U57655 Human dihyd
3	488.2	56.7	854	93	X95670 H.sapiens D
4	171.2	19.9	3951	97	U09178 Human dihyd
5	171.2	19.9	3957	9	AR027740 Sequence
6	169.6	19.7	3078	85	AB003063 Homo sapi
7	169.6	19.7	4409	97	U20938 Human lymph
8	153.4	17.8	4421	7	BTU20981 Bos taurus

```

9 150.2 17.4 4441 7 SSU09179
10 150.2 17.4 4447 9 AR027741
11 145.2 16.9 4358 94 D85035
12 70.6 8.2 43803 6 CELC25F6
13 70 8.1 24429 63 AC014157
14 70 8.1 303092 4 AE003446
15 68.4 7.9 2348 6 DM055491
16 55.4 6.4 117674 84 DMR330C13
17 51 5.9 141953 71 AC044793
18 50 5.8 156060 60 AC004153
19 49.8 5.8 150803 61 AC010228
20 49.6 5.8 1017 6 DDCPR020
21 49.4 5.7 147490 90 AL445207
22 49.4 5.7 160728 70 AC026816
23 49.4 5.7 161025 66 AC021276
24 49.4 5.7 164450 82 AP000794
25 49.2 5.7 168674 69 AC025645
26 48.8 5.7 94536 15 T14F8
27 48.8 5.7 194892 13 ATCHRIV6
28 48.8 5.7 217327 79 AL162723
29 48.2 5.6 1149 5 AF044859
30 48.2 5.6 12029 4 AE001412
31 48 5.6 590 96 PARMTD14D
32 48 5.6 692 96 PARMTD14H
33 48 5.6 726 96 PARMTD14C
34 48 5.6 760 96 PARMTD14A
35 48 5.6 40469 6 MIPAGEN
36 48 5.6 47573 5 AF030694
37 47.4 5.5 112519 85 AB041992
38 47.4 5.5 148965 72 AC060784
39 47.4 5.5 165823 71 AC027625
40 47.4 5.5 177920 64 AC016700
41 47.4 5.5 300000 91 AP002529
42 47.2 5.5 86827 96 FEMAL3P5
43 47 5.5 7218 10 I66494
44 47 5.5 65691 96 FEMAL3P1
45 46.8 5.4 87509 12 AC004482

```

ALIGNMENTS

```

RESULT 1
AL356574/c
LOCUS AL356574 180286 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-359C24, *** SEQUENCING IN
PROGRESS ***, 26 unordered pieces.
ACCESSION AL356574
VERSION AL356574.3 GI:9797568
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 180286)
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9213933.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA359C24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 166105 bases at least Q40

```

FEATURES source

```

Consensus quality: 171672 bases at least Q30
Consensus quality: 174563 bases at least Q20
Insert size: 177786; sum-of-contigs
Insert size: 186027; 5.1% error; agarose-fp
Quality coverage: 3.18x in Q20 bases; sum-of-contigs Quality
coverage: 3.10x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15090: contig of 15090 bp in length
* 15091 15190: gap of 100 bp
* 15191 23545: contig of 8355 bp in length
* 23546 23645: gap of 100 bp
* 23646 26766: contig of 3121 bp in length
* 26767 26866: gap of 100 bp
* 26867 38453: contig of 11587 bp in length
* 38454 38553: gap of 100 bp
* 38554 45878: contig of 7325 bp in length
* 45879 45978: gap of 100 bp
* 45979 51984: contig of 6006 bp in length
* 51985 52084: gap of 100 bp
* 52085 58204: contig of 6120 bp in length
* 58205 58304: gap of 100 bp
* 58305 71630: contig of 13326 bp in length
* 71631 71730: gap of 100 bp
* 71731 75338: contig of 3608 bp in length
* 75339 75438: gap of 100 bp
* 75439 85213: contig of 9775 bp in length
* 85214 85313: gap of 100 bp
* 85314 88517: contig of 3204 bp in length
* 88518 92454: contig of 3837 bp in length
* 92455 92554: gap of 100 bp
* 92555 98200: contig of 5646 bp in length
* 98201 98300: gap of 100 bp
* 98301 111029: contig of 12729 bp in length
* 111030 111129: gap of 100 bp
* 111130 117311: contig of 6182 bp in length
* 117312 117411: gap of 100 bp
* 117412 121093: contig of 3682 bp in length
* 121094 121193: gap of 100 bp
* 121194 137336: contig of 16143 bp in length
* 137337 137436: gap of 100 bp
* 137437 141486: contig of 4050 bp in length
* 141487 141586: gap of 100 bp
* 141587 146703: contig of 5117 bp in length
* 146704 146803: gap of 100 bp
* 146804 149034: contig of 2231 bp in length
* 149035 149134: gap of 100 bp
* 149135 154764: contig of 5630 bp in length
* 154765 154864: gap of 100 bp
* 154865 159990: contig of 5126 bp in length
* 159991 160090: gap of 100 bp
* 160091 163742: contig of 3652 bp in length
* 163743 163842: gap of 100 bp
* 163843 166322: contig of 2480 bp in length
* 166323 166422: gap of 100 bp
* 166423 176700: contig of 10278 bp in length
* 176701 176800: gap of 100 bp
* 176801 180286: contig of 3486 bp in length.
Location/Qualifiers
1. 180286
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-359C24"
/clone_lib="RPC1-11.2"

```

```
misc_feature 1. .15090
/note="assembly_fragment:00687
clone_end:SP6
vector_side:left"
15191. .23545
/note="assembly_fragment:00038
fragment_chain:1"
23646. .26766
/note="assembly_fragment:00972
fragment_chain:1"
26867. .38453
/note="assembly_fragment:01030
fragment_chain:1"
38554. .45878
/note="assembly_fragment:00178
fragment_chain:2"
45979. .51984
/note="assembly_fragment:00498
fragment_chain:2"
52085. .58204
/note="assembly_fragment:00217
fragment_chain:3"
58305. .71630
/note="assembly_fragment:01029
fragment_chain:3"
71731. .75338
/note="assembly_fragment:00265
fragment_chain:4"
75439. .85213
/note="assembly_fragment:01189
fragment_chain:4"
85314. .88517
/note="assembly_fragment:00320
fragment_chain:5"
88618. .92454
/note="assembly_fragment:00693
fragment_chain:5"
92555. .98200
/note="assembly_fragment:00834
fragment_chain:6"
98301. .111029
/note="assembly_fragment:00385
fragment_chain:6"
111130. .117311
/note="assembly_fragment:00932
fragment_chain:7"
117412. .121093
/note="assembly_fragment:01129
fragment_chain:7"
121194. .137336
/note="assembly_fragment:00160"
137437. .141486
/note="assembly_fragment:00194"
141587. .146703
/note="assembly_fragment:00234"
146804. .149034
/note="assembly_fragment:00529"
149135. .154764
/note="assembly_fragment:00590"
154865. .159990
/note="assembly_fragment:01001"
160091. .163742
/note="assembly_fragment:01119"
163843. .166322
/note="assembly_fragment:01294"
166423. .176700
/note="assembly_fragment:01323"
176801. .180286
/note="assembly_fragment:01326"
BASE COUNT 57303 a 32087 c 31562 g 56812 t 2522 others
ORIGIN
```

```
Query Match 90.0%; Score 775.2; DB 80; Length 180286;
Best Local Similarity 95.8%; Pred. No. 2.le-142;
Matches 828; Conservative 0; Mismatches 33; Indels 3; Gaps 3;

QY 1 TGTAAATGAAGATAAATATTTTGGTTTTTTCGCTGTTCTAAACCTAGGTTTACAGAACT 60
|||||
DB 50426 TGTAAATGAATATAATATTTTATTTTTCGCTATTTCTAAACCTAGAAATTAACAAGACT 50367
|||||

QY 61 AATTTATCTGGAGCTAACAAATACTTTTATTTTACCTTTTATTTTGCAGTAGTTATGTT 120
|||||
DB 50366 AATTTAGCTAAAGCTAACAAATACTTTTATTTTACCTTTTATTTTGCAGTAGTTATGTT 50307
|||||

QY 121 CAATTTCAATTTAATGTATATTTAAATAATTCCTCTGCAATATGTGAGGAGGACCTCAT 180
|||||
DB 50306 CAATTTCAATTTAATGTATATTTAAATAATTCCTCTGCAATATGTGAGGAGGACCTCAT 50247
|||||

QY 181 AATATTGTCTATATGGAATGAGCAGATAATAAGATTATAGCTTTTCTTGTCAAAGG 240
|||||
DB 50246 AATATTGTCTATATGGAATGAGCAGATAATAAGATTATAGCTTTTCTTGTCAAAGG 50187
|||||

QY 241 AGACTCAATATCTTTACTCTTTTCATGAGGACATTTGTGACAAATGTTCCCCCATATCAT 300
|||||
DB 50186 AGACTCAATATCTTTACTCTTTTCATGAGGACATTTGTGACAAATGTTCCCCCATATCAT 50127
|||||

QY 301 CCGGGGAACCACTCTGGCCCATCTATGGCCCTGGACAAAAGCTCTTCTGAATATTGA 360
|||||
DB 50126 CCGGGGAACCACTCTGGCCCATCTATGGCCCTGGACAAAAGCTCTTCTGAATATTGA 50067
|||||

QY 361 GCTCATCAGTGAGAAACGGCTGCATATTTGGTGTCAAAGTGTCACTGAACCTAAAGCTGA 420
|||||
DB 50066 GCTCATCAGTGAGAAACGGCTGCATATTTGGTGTCAAAGTGTCACTGAACCTAAAGCTGA 50007
|||||

QY 421 CTTCCAGACAACGTAAGTGTGA-TAAAACTTAAACAAGAGAATTTGGCATAAGTTGGT 479
|||||
DB 50006 CTTCCAGACAACGTAAGTGTGA-TAAAACTTAAACAAGAGAATTTGGCATAAGTTGGT 49947
|||||

QY 480 GAATGTTTATTTAAACATCCAAATTCATAGGCTTATAAATATTTAATCTGTATATTTATCA 539
|||||
DB 49946 GAATGTTTATTTAAACATCCAAATTCATAGGCTTATAAATATTTAATCTGTATATTTATTA 49887
|||||

QY 540 AGCAATCTGCCAGTTGCTTTGCTGATGATAGCAAGAGATAAAAAAGAAAAGCTCAAG 599
|||||
DB 49886 AGCAATCTGCCAGTTGCTTTGCTGATGATAGCAAGAGATAAAAAAGAAAAGCTCAAG 49827
|||||

QY 600 AACTCATAAAAAACCCACACAATGTGAAGCTCTGTTTATAAATGGTGCCTATGAAGATGA 659
|||||
DB 49826 AACTCATAAAAAACCCACACAATGTGAAGCTCTGTTTATAAATGGTGCCTATGAAGATGA 49767
|||||

QY 660 AGAAGTATCTACATAAGCAGAAGAGAAATGAATACTACTATTTTATGTGTTGGCCC 719
|||||
DB 49766 AGAAGTATCTACATAAGCAGAAGAGAAATGAATACTACTATTTTATGTGTTGGTTT 49707
|||||

QY 720 CCACGTATGCGGTGGTATTTTATGAAGTGATGACC-CAGGAAGAAATTTGAACACTATA 778
|||||
DB 49706 TCACGTATGCGGTGGTATTTTATGAAGTGATGACCCTAGGAAGAAATTTGAACACTATA 49647
|||||

QY 779 AACCACTCAAAATATAAACCAGGAGCAGACGATATC-TCCTATGAAGCTCTGATTT 837
|||||
DB 49646 AATCAATTTAAATATAAACCTGAGGCAAGACGATATCTTCTTATGAAGTCTATATTT 49587
|||||

QY 838 ACTCAGTGGGAATAATTTATTAA 861
|||||
DB 49586 TTTCACTGGGAATAATTTATTAA 49563
|||||

RESULT 2
HSU57655
LOCUS HSU57655 626 bp DNA PRI 01-SEP-1996
DEFINITION Human dihydropyrimidine dehydrogenase (DPYD) gene, partial cds.
ACCESSION U57655
VERSION U57655.1 GI:1518606
KEYWORDS human.
```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS Fernandez-Salguero, P.M., Wei, X., Jones, S., Idle, J.R. and Gonzalez, F.J.
TITLE Lack of dihydropyrimidine dehydrogenase activity and thymine-uraciluria caused by a common splice mutation Hum. Mutat. (1996) In press
REFERENCE 2 (bases 1 to 626)
AUTHORS Fernandez-Salguero, P.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-1996) Pedro M. Fernandez-Salguero, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES
Source 1. 626
/organism="Homo sapiens"
/db_xref="taxon:9606"
179. 343
/gene="DPYD"
/gene="DPYD"
/function="uracil and thymine metabolism"
/codon_start=1
/evidence=experimental
/product="dihydropyrimidine dehydrogenase"
/protein_id="AAB07049.1"
/db_xref="GI:1518607"
/translation="DIVTNVSPRIIRGTSGPMYGPQSSFLNIELISEKTAAYWCQS VTEUKADFPDN"
BASE COUNT 221 a 100 c 117 g 188 t
ORIGIN

Query Match 70.7%; Score 608.6; DB 97; Length 626;
Best Local Similarity 99.2%; Pred. No. 9.2e-110;
Matches 622; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 90 TTTACCTTTTATTTCAAGTAGTTATGTTCAATTCCTAATTTAATGATATTAATAAATT 149
DB 1 TTTACCTTTTATTTCAAGTAGTTATGTTCAATTCCTAATTTAATGATATTAATAAATT 60

QY 150 CCTCTGCAATATGTGAGGAGGACCTCATATAAATATTGTCTATATGGAATGAGCAGATA 209
DB 61 CCTCTGCAATATGTGAGGAGGACCTCAT-AAAATTGTCTATATGGAATGAGCAGATA 119

QY 210 ATAAAGATTATAGCTTTCTTTGTCAAAGGAGACTCAATATCTTTACTCTTTCATGAGG 269
DB 120 ATAAAGATTATAGCTTTCTTTGTCAAAGGAGACTCAATATCTTTACTCTTTCATCAGG 179

QY 270 ACATTGTGACAAATGTTCCGCCAATAATCATCCGGGGAACCACTCTGCGCCCATGTATG 329
DB 180 ACATTGTGACAAATGTTCCGCCAATAATCATCCGGGGAACCACTCTGCGCCCATGTATG 239

QY 330 GCCCTGGACAAAGCTCCTTTCTGTAATTTGAGCTCATCAGTGAGAAAACGGCTGCATATT 389
DB 240 GCCCTGGACAAAGCTCCTTTCTGTAATTTGAGCTCATCAGTGAGAAAACGGCTGCATATT 299

QY 390 GGTGTCAAAGTGTCACTGAACCTAAAGGCTGACTTCCAGACACACGTAAGTGTGATAAAA 449
DB 300 GGTGTCAAAGTGTCACTGAACCTAAAGGCTGACTTCCAGACACACGTAAGTGTGATAAAA 359

QY 450 TCTAAACAGAGAATTGGCATAAGTTGGTGAATGTTTATTAACATCCCAATTCATAGG 509
DB 360 TCTAAACAGAGAATTGGCATAAGTTGGTGAATGTTTATTAACATCCCAATTCATAGG 419

QY 510 CTTATAAATATTATGTGTATATTTTATCAAGAACTCTGCCAGTTGCTTTTCTGATGCGAT 569
DB 420 CTTATAAATATTATGTGTATATTTTATCAAGAACTCTGCCAGTTGCTTTTCTGATGCGAT 479

QY 570 AGAAGATATATAAAGACAAAGAGCTCAAGACACTCAATAAAACCCACACATGTGAAGCT 629
DB 480 AGAAGATATATAAAGACAAAGAGCTCAAGAACTCAATAAAACCCACACATGTGAAGCT 539

QY 630 CTGTTATAAATGGTGCCCATGTAGATGAGAGAGTATCTACATAGCAGAGGAAGAGA 689
DB 540 CTGTTATAAATGGTGCCCATGTAGATGAGAGAGTATCTACATAGTAGAGGAAGAGA 599

QY 690 AATGAATACTCAATTTTATGAGTTGG 716
DB 600 AATGAATACTCAATTTTATGAGTTGG 626

RESULT 3
HSDYDPGEN HSDYDPGEN 854 bp DNA PRI 17-FEB-1997
LOCUS H.sapiens DPYD gene, partial sequence.
DEFINITION X95670
ACCESSION X95670.1 GI:1246752
VERSION dihydropyrimidine dehydrogenase; DPYD gene; exon X; mutation.
KEYWORDS SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 854)
AUTHORS Vreken, P., Van Kuilenburg, A.B., Meinsma, R., Smit, G.P., Bakker, H.D., De Abreu, R.A. and van Gennip, A.H.
TITLE A point mutation in an invariant splice donor site leads to exon skipping in two unrelated Dutch patients with dihydropyrimidine dehydrogenase deficiency
J. Inherit. Metab. Dis. 19 (5), 645-654 (1996)
JOURNAL 97047101
MEDLINE 2 (bases 1 to 854)
REFERENCE Vreken, P.
AUTHORS Direct Submission
TITLE Submitted (14-FEB-1996) P. Vreken, Academic Med.Center, Univ. of Amsterdam, Dep. of Pediatrics and Clinical Chemistry, F0-224, P.O. Box 22700, NL-1100 DE Amsterdam, NETHERLANDS
JOURNAL Related sequences:-
COMMENT Meinsma et al., DNA Cell Biol. 14:1-6 (1995), U20938, and Yokota H. et al., J. Biol. Chem. 269:23192-23196 (1994).

FEATURES
source 1. 854
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="liver"
/map="p22"
<1. 568
569. 734
/gene="DPYD"
<569. 734
/gene="DPYD"
/EC_number="1.3.1.2"
/note="exon X"
/codon_start=2
/product="dihydropyrimidine dehydrogenase (NADP+)"
/protein_id="CAA64973.1"
/db_xref="GI:1246753"
/db_xref="SWISS-PROT:Q12882"
/translation="DIVTNVSPRIIRGTSGPMYGPQSSFLNIELISEKTAAYWCQS VTEUKADFPDN"
569. 734
/gene="DPYD"
725
/gene="DPYD"
/replace="c"
735. 7854
735
/note="changes the splice donor sequence and leads to exon skipping"
/replace="a"
BASE COUNT 301 a 126 c 130 g 297 t
ORIGIN

Query Match	56.7%	Score 488.2;	DB 93;	Length 854;
Best Local Similarity	96.2%	Pred. No. 3.5e-86;		
Matches	532;	Conservative	0;	Mismatches 18; Indels 3; Gaps 3;
QY	1	TGTTAATGAGATAAATATTTTCTTTTTCGCTGTTCTAAACCTAGGTTTACAGAAGT	60	
Db	304	TGTTAATGAGATAAATATTTTATTTTGGCTATTCTTAACCTAGAAATTACAGAAGT	363	
QY	61	AATTATCTGAGCTAAACAATCTTTATTTTACCTTTTATTTTTCGCAAGTACTTTATGTT	120	
Db	364	AATTAGCTAAGCTAAACAATCTTTATTTTACCTTTTATTTTTCGCAAGTACTTTATGTT	422	
QY	121	CAATCTTAATTAATGATATATAAATAATCTCTGCAAAATGAGAGGACCTCAT	180	
Db	423	CAATCTTAATTAATGATATATAAATAATCTCTGCAAAATGAGAGGACCTCAT	482	
QY	181	AAATATGTCATATGGAATGAGAGCAGATAAATAAGATTATAGCTTTTCTTGTCAAAAGG	240	
Db	483	AAATA-TGTCATATGGAATGAGAGCAGATAAATAAGATTATAGCTTTTCTTGTCAAAAGG	541	
QY	241	AGACTCAATATCTTTACTCTTTTCATGAGGACATTTGTGACAAATGTTTCCGCCCAATCAT	300	
Db	542	AGACTCAATATCTTTACTCTTTTCATGAGGACATTTGTGACAAATGTTTCCGCCCAATCAT	601	
QY	301	CCGGGGAACCACTCTGCCCCCATGATGCGCCCTGGACAAAGCTTCTTCTGCAATATGGA	360	
Db	602	CCGGGGAACCACTCTGCCCCCATGATGCGCCCTGGACAAAGCTTCTTCTGCAATATGGA	661	
QY	361	GCTCATCAGTGAGAAACGGCTGCATATTTGGTGCAAGTGTCACTGAACCTAAAGGCTGA	420	
Db	662	GCTCATCAGTGAGAAACGGCTGCATATTTGGTGCAAGTGTCACTGAACCTAAAGGCTGA	721	
QY	421	CTTCCAGACAACCTAAGTGTGA-TAAAAATCTAAAAACAAGAGAAATGGCATAAGTTGGT	479	
Db	722	CTTCCAGACAACCTAAGTGTGA-TTACATCTAAACAAGAGAAATGGCATAAGTTGGT	781	
QY	480	GAATGTTTATTAACATCAATTCATAGGCTTATAAATATTAATGTTATATTTATCA	539	
Db	782	GAATGTTTATTAACATCAATTCATAGGCTTATAAATATTAATGTTATATTTATTA	841	
QY	540	ACGAATCTGCCAG	552	
Db	842	AGAAATCTGCCAG	854	
RESULT	4			
HSU09178				
LOCUS	HSU09178	3951 bp	mRNA	PRI 28-DEC-1994
DEFINITION	Human dihydropyrimidine dehydrogenase mRNA, complete cds.			
ACCESSION	U09178			
VERSION	U09178.1	GI:558304		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 3951)			
AUTHORS	Yokota, H., Fernandez-Salguero, P., Furuya, H., Lin, K., McBride, O.W., Podschun, B., Schnackerz, K.D. and Gonzalez, F.J.			
TITLE	cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and congenital thymine uraciluria			
JOURNAL	J. Biol. Chem. 269 (37), 23192-23196 (1994)			
MEDLINE	94365020			
REFERENCE	2 (sites)			
AUTHORS	Eggink, G., Engel, H., Vriend, G., Terpstra, P. and Witholt, B.			
TITLE	Rubredoxin reductase of Pseudomonas oleovorans. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints			
JOURNAL	J. Mol. Biol. 212 (1), 135-142 (1990)			
MEDLINE	90204534			
REFERENCE	3 (sites)			
AUTHORS	Porter, D.J., Chestnut, W.G., Merrill, B.M. and Spector, T.			
QY	244	CTCAATATCTTTACTCTTTTATGAGGACATTTGTGACAAATGTTTCCCAATATCATCCG	303	
Db	1797	CACCAAACTTTCTCTTGTATAGGACATTTGTGACAAATGTTTCCCAATATCATCCG	1856	
Query Match	19.9%	Score 171.2;	DB 97;	Length 3951;
Best Local Similarity	93.2%	Pred. No. 4.2e-24;		
Matches	179;	Conservative	0;	Mismatches 13; Indels 0; Gaps 0;
BASE COUNT	1153 a	785 c	896 g	1117 t
ORIGIN				
1	3951			
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/tissue_type="liver"				
82.3159				
/codon_start=1				
/product="dihydropyrimidine dehydrogenase"				
/protein_id="AAA57474.1"				
/db_xref="GI:558305"				
/translation="MAPVLSKDSADIESILALNPRTQTHATLCTSAKLDKKHKRN				
PDKNCFNCEKLENNFDDIKHTLGERGALREAMRLCKADAPCKQKSCPTNLDIKSFIT				
SIANKNYGAACKMIFSDNPLGLTCGMVCPSTSLCVGCNLYATEEGPINIGGLQQRAT				
EYFKAMSIQIRNPSPPPKMEASAKIALFCGAPASISCASFLARLGYSIDTIFE				
KQYVGLSTSEIPQRPDYVNVFEIEMKDLGVKLIICKLSLVNEMTUSTLEKGY				
KAAFIGILPEPNKDAIFOGLTQDQGFYTSKDFLPLVAKSGKAGMACHEKCEPLFS				
VIVLGAGDTAFDCATSAIRCGARRVIFVRKGFVNIRAVPEEMELAKCEKCEPLFS				
PRKVIKGGRIVAMQFRTQDETQKWNEDQDQVHLKADVTISAFSGVLSLDPKVEA				
LSPIKFNWGLPEVPTMOTSEAWPVAGDVGLANTVTVESVNDCKQASWYTHKYVQ				
SOYGASYSKAPPELPETPIDLDVDSVEMAGLAFINPFGLASATPATSTSMIRAREA				
GWGFALTFTSLDKDIVNSPRIIRGTSGPMVPGQSFNLIELISAKTAAYWCQS				
VTBKADFPDNIIVIASIMCYNKNDWTSLAKSDSGDALELNLSLCPHGMGRMGL				
ACQDPELVNRCVWRVQAVQIPFFAKLTNPVTDIVISARAAGGANGVTATNTVSG				
LMGLKSDGTPWPAVGLAKRTTYGGVSGCTAIRPIALRAVTSIARALPGFPIATGSDS				
AESEQLFHSGASVQVCSAIONQDVTIEDYCTGLKALLYLSIELODWDGOSPAT				
VSHQKGPVRIAEMLDKLPSFGLPEQRKKIIAENKIRLKEQNVAFLPKSCFIP				
FDPETIKDVGKALQYLGTFGELSNEVQVAMIDEMCINCKYKSCFIP				
FDPTHLPTITDTCCTCLSLVCPIDVCIKMYSRTPPEYKRGVPLSVNPVC"				
1084.1134				
/citation=[5]				
/function="catalytic cofactor NADPH/NADP binding site"				
1492.1524				
/citation=[2]				
/function="electron transfer center, FAD binding site"				
2062.2175				
/citation=[3]				
/function="uracil (substrate) binding site"				
2938.2973				
/standard_name="iron-sulfur center"				
/citation=[4]				
/function="catalytic cofactor [4Fe-4S] binding site"				

QY 304 GGAACCACTCTGGCCCATGATGGCCCTGGACAAAGCTCTTTCTGAATATTGAGCT 363
|||||
Db 1857 GGAACCACTCTGGCCCATGATGGCCCTGGACAAAGCTCTTTCTGAATATTGAGCT 1916
|||||
QY 364 CATCAGTGAGAAACGGCTGCATATTGTTGTCAAAAGTGTCACTGAAGCTGACTT 423
|||||
Db 1917 CATCAGTGAGAAACGGCTGCATATTGTTGTCAAAAGTGTCACTGAAGCTGACTT 1976
|||||
QY 424 CCCAGACAACGT 435
|||||
Db 1977 CCCAGACAACAT 1988
|||||
RESULT 5
AR027740 3957 bp DNA PAT 29-SEP-1999
LOCUS AR027740 Sequence 1 from patent US 5856454.
DEFINITION AR027740
ACCESSION AR027740
VERSION AR027740.1 GI:5938560
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
UNCLASSIFIED.
REFERENCE 1 (bases 1 to 3957)
AUTHORS Gonzalez,F.J. and Fernandez-Salguero,P.
TITLE CDNA for human and pig dihydropyrimidine dehydrogenase
JOURNAL Patent: US 5856454-A 1 05-JAN-1999;
LOCATION/Qualifiers
FEATURES
source 1..3957
/organism="unknown"
BASE COUNT 1156 a 787 c 897 g 1117 t
ORIGIN
Query Match 19.98; Score 171.2; DB 9; Length 3957;
Best Local Similarity 93.28; Pred. No. 4.2e-24;
Matches 179; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 244 CTCATATCTTACTCTTTCATGAGGACATGTGACAAATGTTTCCCCCAATAATCATCCG 303
| | | | |
Db 1803 CACCAAACTTCTCTTGTATAGGACATGTGACAAATGTTTCCCCCAATAATCATCCG 1862
| | | | |
QY 304 GGAACCACTCTGGCCCATGATGGCCCTGGACAAAGCTCTTTCTGAATATTGAGCT 363
|||||
Db 1863 GGAACCACTCTGGCCCATGATGGCCCTGGACAAAGCTCTTTCTGAATATTGAGCT 1922
|||||
QY 364 CATCAGTGAGAAACGGCTGCATATTGTTGTCAAAAGTGTCACTGAAGCTGACTT 423
|||||
Db 1923 CATCAGTGAGAAACGGCTGCATATTGTTGTCAAAAGTGTCACTGAAGCTGACTT 1982
|||||
QY 424 CCCAGACAACGT 435
|||||
Db 1983 CCCAGACAACAT 1994
|||||
RESULT 6
AB003063 3078 bp mRNA PRI 21-JAN-2000
LOCUS AB003063 Homo sapiens mRNA for dihydropyrimidine dehydrogenase, complete
DEFINITION cds.
ACCESSION AB003063
VERSION AB003063.1 GI:6729337
KEYWORDS dihydropyrimidine dehydrogenase.
SOURCE Homo sapiens male lymphocyte cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Ojura,K., Nishiyama,T., Takubo,H., Kato,A., Okuda,H., Arakawa,K.,
Fukushima,M., Nagayama,S., Kawaguchi,Y. and Watabe,T.
TITLE Suicidal inactivation of human dihydropyrimidine dehydrogenase by
(E)-5-(2-bromovinyl)uracil derived from the antiviral, sorivudine

JOURNAL Cancer Lett. 122 (1-2), 107-113 (1998)
MEDLINE 98124145
REMARK Erratum: [[published erratum appears in Cancer Lett 1998 Jun
19;128(2):229]]
REFERENCE 2 (bases 1 to 3078)
AUTHORS Ojura,K.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1997) to the DDBJ/EMBL/GenBank databases.
Kenichiro Ogura, Tokyo University of Pharmacy and Life Science,
Department of Drug Metabolism and Molecular Toxicology; 1432-1
Horinouchi, Hachioji, Tokyo 192-0392, Japan
(E-mail:ogurake@toyaku.ac.jp, Tel:+81-426-76-4518,
Fax:+81-426-76-4517)
FEATURES
Location/Qualifiers
source 1..3078
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="lymphocyte"
/sex="male"
1..3078
/codon_start=1
/product="dihydropyrimidine dehydrogenase"
/protein_id="BAA89789.1"
/db_xref="GI:6729338"
/translation="MAPVLSKDSADIESILALNPRTQTHTALCTSAKKLKKHKWRN
PKNCFNCEKLENNFDDIKHTLTGERGALREARMLKCADAPQKSCPTNLDIKSFIT
SIANKNYGAAKMIFSDNPLGLTCGMVPTSDLCVGCNLYATVEGFINIGLQOAFIT
EYFKAMSIQIRNPSLPPEPKMSEAYSAKIALFCAGPASTSCASFLARLGYSDITIE
KQEVGGLSTSEIPQLPYDVVNFELMKDLGVKIIKCGSKLSVNMETLSTLKEGY
KAAFIGILPEPNKDAIFOGLTQDQGYTSLPLVAKGSKAGKACHSPLSIRGV
VIVLGADTAFDCATSAIRGARRVIVFRKGFVNTRAVPEEMELAKEKECEFLPFLS
PRKIVKGRIVAMQFVTEQDETKWNEDEDMVHLKADVITSAFSGVSLSDPKVKA
LSPIKENRWGLPEVDPTMOTSEAVFAGGVGLANTTVESVNDGKQASWY IHKYVQ
SOYGASVSAPKPELPLEYTPIDLVDISVEMAGLAFINPGLASATPATSTWIRAFEA
GNGFALTFTSLDKDIVTVNSPRIIRGTTSGPMYGCQSSEFLNIELISEKTAAYWCS
VTELKADPDNIIVIASIMCSYNKNDWTELAKSDESDGADELNLSCPHGMGERGMGL
ACQDDELVRNICRWVRVQVPIFFAKLTNPVTDIVISARAKEGGANGATNTVSG
LMGLKSDGTWPVAVGIAKRTTYGVSQGTAPRIALRAVTSIARALGFPPLATGGIDS
VSHQKGPVPRIAEMLMDKLLPSRGPYLEQRKKIAENKIRLKEQNVAFSPLKRNCFIP
KPIPTIKDVIKALQYLTGCELSNVEOVVAMIDEMCINCCGKYMTCDNSGVQAIQ
FDPETHLPITDTCCTGCTCLCSVCPVDCIKMVSRTTTPYEPKRGVPLSVNPVC"
BASE COUNT 875 a 643 c 734 g 826 t
ORIGIN
Query Match 19.7%; Score 169.6; DB 85; Length 3078;
Best Local Similarity 92.7%; Pred. No. 8.6e-24;
Matches 178; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 244 CTCATATCTTACTCTTTCATGAGGACATGTGACAAATGTTTCCCCCAATAATCATCCG 303
| | | | |
Db 1716 CACCAAACTTCTCTCTTGTATAGGACATGTGACAAATGTTTCCCCCAATAATCATCCG 1775
| | | | |
QY 304 GGAACCACTCTGGCCCATGATGGCCCTGGACAAAGCTCTTTCTGAATATTGAGCT 363
|||||
Db 1776 GGAACCACTCTGGCCCATGATGGCCCTGGACAAAGCTCTTTCTGAATATTGAGCT 1835
|||||
QY 364 CATCAGTGAGAAACGGCTGCATATTGTTGTCAAAAGTGTCACTGAAGCTGACTT 423
|||||
Db 1836 CATCAGTGAGAAACGGCTGCATATTGTTGTCAAAAGTGTCACTGAAGCTGACTT 1895
|||||
QY 424 CCCAGACAACGT 435
|||||
Db 1896 TCCAGACAACAT 1907
|||||
RESULT 7
HSU20938 4409 bp mRNA PRI 04-APR-1997
LOCUS HSU20938 Human lymphocyte dihydropyrimidine dehydrogenase mRNA, complete
DEFINITION cds
ACCESSION U20938

```

VERSION      U20938.1  GI:1926407
KEYWORDS
SOURCE
ORGANISM      human.
               Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 4409)
AUTHORS       Diasio,R.B., Beavers,T.L. and Carpenter,J.T.
TITLE         Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical
               basis for familial pyrimidinemia and severe 5-fluorouracil-induced
               toxicity
JOURNAL       J.Clin. Invest. 81 (1), 47-51 (1988)
MEDLINE       88087908
REFERENCE     2 (bases 1 to 4409)
AUTHORS       Lu,Z.H., Zhang,R. and Diasio,R.B.
TITLE         Purification and characterization of dihydropyrimidine
               dehydrogenase from human liver
JOURNAL       J. Biol. Chem. 267 (24), 17102-17109 (1992)
MEDLINE       92381021
REFERENCE     3 (bases 1 to 4409)
AUTHORS       Johnson,M.R., Albin,N., Shahinian,H. and Diasio,R.B.
TITLE         Identification of a frameshift in the DNA coding for
               dihydropyrimidine dehydrogenase (DPD) in a DPD deficient patient
               exhibiting 5-fluorouracil toxicity
JOURNAL       Unpublished
REFERENCE     4 (bases 1 to 4409)
AUTHORS       Johnson,M.R.
TITLE         Direct Submission
JOURNAL       Submitted (10-FEB-1995) Pharmacology, University of Alabama at
               Birmingham, 1670 University Blvd., Birmingham, AL 35294-0019, USA
               On Apr 4, 1997 this sequence version replaced gi:693911.
               Location/Qualifiers
               1..4409
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /cell_type="Lymphocyte"
               102..3179
               /EC_number="1.2.1.2"
               /note="DPD; dihydrouracil dehydrogenase"
               /codon_start=1
               /product="dihydropyrimidine dehydrogenase"
               /protein_id="AAB51366.1"
               /db_xref="GI:693912"
               /translation="MAPVLISKDVADIESILALNPRTQSTATLSTLAKLDDKKHWRN
               PDKNFCNLENNFDDIKHTLGERGALREARMLCKADAPCKQSCPTNLDLKSFIT
               SIANKNYGAAKMFSDNPLGLTCGMVCPSTSLCVGCNLYATEEGPINTGGLOQAT
               EVFKANSIQIRNPSLPPEKMESEAYSAKIALFGAGPASISCFARLQYSITLIE
               KQYVGLSTSEIQFRLPYDVVNFIEMLKDLGVKILCGKSLSVNEMTLSTLKEGY
               KAAFIGIGLPEPKKDHIFOGLTODQGFYTSKDFPLVAKSKAGMACHSPLPSIRG
               VIVLAGDADFDCATGALRCGARRVIFVRKGFVNIRAVPEEMELAKAEKCEFLPLS
               PRKVIKGRIVAMQFVRTEQDETGWNEDEDMVHLKADVVISAFSGVLSLDPKVEA
               LSPIKENRWGLPEVDPETMOTSEAWFAGDVVGLANTVSVESVNDGKQASWYTHKYVO
               SOYGASVSAKPELPLEYTPIDLVDISVEMAGLKEINPFGLASATPATSTSMIRAFEA
               GWGFALTFTSLDKDITVNSPRIIRGTTSGPMYGPQGSFLNIELISEKTAAYWCQS
               VTELKADFPDNIIVIASIMCSYNKNDWTELAKKSEDSADALENLSCPGHGMGERMGL
               ACQDDELVNRICWRVQVQIPFFAKLTNPVDIVSIARAAGEGANGVATNTVSG
               LMKLGSDGTPWPAVIAKTRTYGVSQTALRPIALRAVTSIARALPGFTILATGSDS
               AESGLQFLHSGASVLOVCSAIONQDETVIDCTGLKALLYLKSIELODWDGQSPAT
               VSHQKGPVPRIAEMLDKLPSFGVLEQRKLLIAENKIRLKEQNVAFSPLKKNCFIP
               KRPIPTIKLQYLGALQYLGELSNEQVAMDEMCINCCKYMTNDSGYQAIQ
               FDEPTHLPTIDFTCTCTCLSLSCVPTDCKIMVSRTPPEPRKGVPLSVNPGV"
BASE COUNT   1317 a 845 c 949 g 1298 t
ORIGIN
Query Match      19.7%; Score 169.6; DB 97; Length 4409;
Best Local Similarity 92.7%; Pred. No. 8.5e-24;
Matches 178; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 244 CTCATATCTTTACTCTTTTCATGAGGACATGTGACAAAATGTTTCCCCCAATATCATCCG 303
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1817 CACCAAACTTTCTCTCTTGAAGGACATGTGACAAAATGTTTCCCCCAATCATCCG 1876

```

```

QY 304 GGAACACCTCTTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 363
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1877 GGAACACCTCTTGGCCCCATGTATGGCCCTGGACAAAGCTCCTTTCTGAATATTGAGCT 1936

QY 364 CATCAGTGAGAAAACGGCTGCATATTGGTGCAAGTGCTCACTGAACATAAGGCTGACTT 423
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1937 CATCAGTGAGAAAACGGCTGCATATTGGTGCAAGTGCTCACTGAACATAAGGCTGACTT 1996

QY 424 CCCAGACAACTGT 435
      | | | | | | | | | |
Db 1997 TCCAGACAACT 2008

RESULT      8
BTU20981
LOCUS       BTU20981 4421 bp mRNA MAM 17-JAN-1997
DEFINITION Bos taurus liver dihydropyrimidine dehydrogenase mRNA, complete
               cds.
ACCESSION   U20981
VERSION     U20981.1 GI:677950
KEYWORDS
SOURCE      cow.
ORGANISM    Bos taurus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
               Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 4421)
AUTHORS     Lu,Z., Zhang,R. and Diasio,R.B.
TITLE       Dihydropyrimidine dehydrogenase activity in human peripheral blood
               mononuclear cells and liver: population characteristics, newly
               identified deficient patients, and clinical implication in
               5-fluorouracil chemotherapy
JOURNAL     Cancer Res. 53 (22), 5433-5438 (1993)
MEDLINE     94036837
REFERENCE   2 (bases 1 to 4421)
AUTHORS     Albin,N., Johnson,M.R. and Diasio,R.B.
TITLE       cDNA cloning of bovine liver dihydropyrimidine dehydrogenase
JOURNAL     DNA Seq. 6 (4), 243-250 (1996)
MEDLINE     97069948
REFERENCE   3 (bases 1 to 4421)
AUTHORS     Johnson,M.R.
TITLE       Direct Submission
JOURNAL     Submitted (13-FEB-1995) Pharmacology/Toxicology, University of
               Alabama at Birmingham, 1670 University Blvd., Birmingham, AL
               35294-0019, USA
               Location/Qualifiers
               1..4421
               /organism="Bos taurus"
               /db_xref="taxon:9913"
               /tissue_type="liver"
               75..3152
               /EC_number="1.3.1.2"
               /note="dihydrothymine dehydrogenase; DPD"
               /codon_start=1
               /product="dihydropyrimidine dehydrogenase"
               /protein_id="AAB40985.1"
               /db_xref="GI:677951"
               /translation="MAPVLISKDVADIESILALNPRTQSTATLSTLAKLDDKKHWRN
               PDKNFCNLENNFDDIKHTLGERGALREARMLCKADAPCKQSCPTNLDLKSFIT
               SIANKNYGAAKMFSDNPLGLTCGMVCPSTSLCVGCNLYATEEGPINTGGLOQAT
               EVFKANSIQIRNPSLPPEKMESEAYSAKIALFGAGPASISCFARLQYSITLIE
               KQYVGLSTSEIQFRLPYDVVNFIEMLKDLGVKILCGKSLSVNEMTLSTLKEGY
               KAAFIGIGLPEPKKDHIFOGLTODQGFYTSKDFPLVAKSKAGMACHSPLPSIRG
               VIVLAGDADFDCATGALRCGARRVIFVRKGFVNIRAVPEELAREKCEFLPLS
               PRKVIKGRIVAMQFVRTEQDETGWNEDEDMVHLKADVVISAFSGVLSLDPKVEA
               LSPIKENRWGLPEVDPETMOTSEAWFAGDVVGLANTVSVESVNDGKQASWYTHKYVO
               SOYGASVSAKPELPLEYTPIDLVDISVEMAGLKEINPFGLASATPATSTSMIRAFEA
               GWGFALTFTSLDKDITVNSPRIIRGTTSGPMYGPQGSFLNIELISEKTAAYWCQS
               VTELKADFPDNIIVIASIMCSYNKNDWTELAKKSEDSADALENLSCPGHGMGERMGL
               ACQDDELVNRICWRVQVQIPFFAKLTNPVDIVSIARAAGEGANGVATNTVSG
               LMKLGSDGTPWPAVIAKTRTYGVSQTALRPIALRAVTSIARALPGFTILATGSDS
               AESGLQFLHSGASVLOVCSAIONQDETVIDCTGLKALLYLKSIELODWDGQSPAT
               VSHQKGPVPRIAEMLDKLPSFGVLEQRKLLIAENKIRLKEQNVAFSPLKKNCFIP
               KRPIPTIKLQYLGALQYLGELSNEQVAMDEMCINCCKYMTNDSGYQAIQ
               FDEPTHLPTIDFTCTCTCLSLSCVPTDCKIMVSRTPPEPRKGVPLSVNPGV"
               KSHQKGPVPCIAELVGKLPSTGPKLEKCKKIIAEKRLAKENVTVLPERNHPIF

```

KKPIPSKDVIGKALQYLGTYGELNNTVOVAVIDEMCINCGKCYMTNCDSGYQAIQ
FDPETHLPVTDTCTGCTCLSVCPIDCIKMWSRITPYEPKRGPLAVNPVS"

BASE COUNT 1315 a 874 c 958 g 1274 t
ORIGIN

Query Match 17.8%; Score 153.4; DB 7; Length 4421;
Best Local Similarity 88.8%; Pred. No. 1.3e-20;
Matches 166; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 246 CAATATCTTTACTCTTCATGAGGACATGTCACAAATGTTTCCCATATATCATCCGGG 305
Db 1792 CCAAACTTTCTCTCTGATAGGATATAGTCACAAATGTTTCCACCAATATCCGGG 1851
QY 306 GAACACCTCTGGCCCCATGATGCGCCTGGACAAAGCTCTTCTGAATATTGAGCTCA 365
Db 1852 GGACACCTCTGGCCCCATGATGCGCCTGGACAAAGCTCTTCTGAATATTGAGCTCA 1911
QY 366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTGTCACTGAACAAAGGCTGACTTCC 425
Db 1912 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTGTCACTGAACAAAGGCTGACTTCC 1971
QY 426 CAGACAA 432
Db 1972 CAGACAA 1978

RESULT 9
SSU09179
LOCUS SSU09179 4441 bp mRNA MAM 28-DEC-1994
DEFINITION Sus scrofa dihydropyrimidine dehydrogenase mRNA, complete cds.
ACCESSION U09179
VERSION U09179.1 GI:558306
KEYWORDS
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 4441)
Yokota, H., Fernandez-Salguero, P., Furuya, H., Lin, K., McBride, O.W.,
Podschun, B., Schnackerz, K.D. and Gonzalez, F.J.
cDNA cloning and chromosome mapping of human dihydropyrimidine
dehydrogenase, an enzyme associated with 5-fluorouracil toxicity
and congenital thymine uraciluria
J. Biol. Chem. 269 (37), 23192-23196 (1994)
94365020
2 (sites)
Eggink, G., Engel, H., Vriend, G., Terpstra, P. and Witholt, B.
Rubredoxin reductase of Pseudomonas oleovorans. Structural
relationship to other flavoprotein oxidoreductases based on one NAD
and two FAD fingerprints
J. Mol. Biol. 212 (1), 135-142 (1990)
90204534
3 (sites)
Porter, D.J., Chestnut, W.G., Merrill, B.M. and Spector, T.
Mechanism-based inactivation of dihydropyrimidine dehydrogenase by
5-ethynyluracil
J. Biol. Chem. 267 (8), 5236-5242 (1992)
92184771
4 (sites)
Dupuis, A., Skehel, J.M. and Walker, J.E.
A homologue of a nuclear-coded iron-sulfur protein subunit of
bovine mitochondrial complex I is encoded in chloroplast genomes
Biochemistry 30 (11), 2954-2960 (1991)
91175743
5 (sites)
Wierenga, R.K., De Maeyer, M.C.H. and Hol, W.G.J.
Interaction of pyrophosphatase moieties with alpha-helices in
dinucleotide binding proteins
Biochemistry 24, 1346-1357 (1985)
6 (bases 1 to 4441)
Gonzalez, F.
Direct Submission

JOURNAL

Submitted (22-APR-1994) Frank J. Gonzalez, National Institutes of
Health, National Cancer Institute, 9000 Rockville Pike, Bethesda,
MD 20892, USA

FEATURES
source

Location/Qualifiers
1. .4441
/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="liver"
82. .3159
/codon_start=1
/product="dihydropyrimidine dehydrogenase"
/protein_id="AAA57475.1"
/db_xref="GI:558307"

CDS

/translation="MAPVLSKDVADIESILALNPRTQSHAAHSLAKLKDKKHKWRN
PDNCFHCEKLENNFGDIKHTLGERGALREAMRCLCADAPCOKSCPTHLJDIKSFIT
SISKNKYGAAMKMFSDNPLGLTCGMVCPPTSLCVGCGNLYATEEGSINTGGLQOFAS
EVFKANNIPOIRNPLPSQKEMPEAYSAKIALGAGPASISCAISALRGYSIDTIFE
KQYVGLSTSEIPRLPYDVVNFIEMLMDLGVKICGRKSSENEITLNTKEEGY
KAAFIGILPEPKTDDIFQGLTQDQGYTSKDFPLVAKSKACMGACACHSPLSIRGA
VIVLGAGDTAFDCATGARRVFLVRKGFVNIARVPEVEVLAKEEKECEFLPFLS
PRKVIKGGRIVAVQVTRTQDEQKWNEDQIVHLKADVIISAFGSVLRDPKVEA
LSPIKRNWDLPEVDPEMOTSEPMVFPAGGDIYGMANTTVESVNDGKOASHYIKYIQ
AQGSYSAKPELPFTYTPVDLDSVEMAGLAFINPFGLSAAPTSSSIRAFEA
GWGFALTFTSLDKDIVNVPRIVRGTTSGPMYGPQSSFLNIELISEKTAAYWCOS
VTELKADFPDNIIVIASIMCSYNKRNWMLSRKAEASGADALELNSCPHGMRGMGL
ACQDPELVNRCRWVRQAVQIPFFAKLTPTNTDIVSIARAKEGGADGVTATNVSG
LMGLKADGTPWPAVGAGKRTTYGGSTAIRPIALRAVTTIARALPGFPPIATGGIDS
AESGLQFLHSGASVLOVCSAVONDETVIODYCTGLKALLYLAKSIEELQWDSQSPGT
ESHQKGPVRIAEMLKGLPNEGPYLEQRKKLIAEEMKRLKQONAAFPPLERKFPFIQ
KLPPIAKIDVIGKALQYLGFTGELSNIQVAVIADIEEMCINCGKCYMTNCDSGYQAIQ
FDPETHLPVTDTCTGCTCLSVCPIDCIKMWSRITPYEPKRGPLAVNPVC"
1084. .1134
/citation=[5]
/function="catalytic cofactor NADPH/NADP binding site"
1492. .1524
/citation=[2]
/function="electron transfer center, FAD binding site"
2052. .2175
/citation=[3]
/function="uracil (substrate) binding site"
2938. .2973
/standard_name="iron-sulfur center"
/citation=[4]
/function="catalytic cofactor [4Fe-4S] binding site"
BASE COUNT 1281 a 928 c 998 g 1234 t
ORIGIN

Query Match 17.4%; Score 150.2; DB 7; Length 4441;
Best Local Similarity 87.7%; Pred. No. 5.4e-20;
Matches 164; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 246 CAATATCTTTACTCTTCATGAGGACATGTCACAAATGTTTCCCATATATCATCCGGG 305
Db 1799 CCAAACTTTCTCTCTGATAGGACATAGTCACAAATGTTTCCACCAAGATCGTCCGGG 1858
QY 306 GAACACCTCTGGCCCCATGATGCGCCTGGACAAAGCTCTTCTGAATATTGAGCTCA 365
Db 1859 GGACTACCTCTGGCCCCATGATGCGCCTGGACAAAGCTCTTCTGAATATTGAGCTCA 1918
QY 366 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTGTCACTGAACAAAGGCTGACTTCC 425
Db 1919 TCAGTGAGAAAACGGCTGCATATTGGTGTCAAAAGTGTCACTGAACAAAGGCTGACTTTC 1978
QY 426 CAGACAA 432
Db 1979 CAGACAA 1985

RESULT 10

AR027741
LOCUS AR027741 4447 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5856454.

[illegible]

Saunders, D., Shownkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstein, L., Wilkinson-Sproat, J., and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

TITLE

elegans

Nature 368 (6466), 32-38 (1994)

94150718

2 (bases 1 to 43803)

Bentley, D.

The sequence of C. elegans cosmid C25F6

Unpublished (1995)

3 (bases 1 to 43803)

Waterston, R.

Direct Submission

Submitted (30-OCT-1995) Robert Waterston

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University,

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

e-mail: rwenan@wustl.edu and jes@sanger.ac.uk

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is C41A3, 1000 bp overlap; 3' end lies in a gap

followed by the cosmid T23F2. Actual start of this cosmid is at

base position 1 of CELC25F6. Actual end is at base position 43803

of CELC25F6.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source

1. 43803

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/map="X"

complement(1397..1495)

/note="Leu; codon recognized: CTA"

complement(5095..5656)

/gene="C25F6.6"

complement(join(5095..5215,5259..5385,5443..5535,

5584..5656))

/gene="C25F6.6"

/codon_start=1

/evidence=not_experimental

/protein_id="AA080431.1"

/db_xref="GI:1049456"

/translation="MMISRFSLKILIVSLIGVLGVADASSSMYARRRTASERIMSTII

RPRMFQPERSSQIKIQLNPALPCCKDESGSICKNLRRDILKLTQKQCTE

PDFSLVCCSCSDAGISYRKRAQKQVFITLVI"

complement(6056..6128)

/note="Lys; codon recognized: AAA"

16243..20895

/gene="C25F6.4"

join(16243..16401,16449..16601,17110..17257,17309..17421,

17576..17675,17720..17825,18125..18202,18252..18358,

18406..18572,18930..19243,19293..19592,20318..20513,

20569..20718,20773..20895)

/gene="C25F6.4"

/note="similar to protein tyrosine kinase"

/codon_start=1

/evidence=not_experimental

/protein_id="AA080432.1"

/db_xref="GI:1049457"

/translation="MQGDIADSQITASSFDKQSGVQPNARLHSELASGAWCPKPKQI

NKSYEFLQVTLNDFLITVSQGRYNGTGRFASHMIDYLRPGSQWIRYKRTG

HYVMDGFDFTTPVIRLDPPIVAVSRNTRTPVPCNTTRVCMRAEIHGCKHGYTST

VPDGRDLTDKDSMEDSQITVETSGIKRGSPANRRFSAASFPEKNQMNWSIG

WNRDTGRTILFPEFVHNFTDVLTAFGRNIDGIDVIFSDGKTFFLFSQISSSE

RQSLNNTSRDYDVPVLRNAGRKVRISIKFSSDMMLTEVHTSAULTLSEKIPP

PSSAATQOLIVGCIITFTFACVAVCVSCLARRQKNKSVDNSVWKKDLITTHMGKPK

TCHVFPNSGNLNGEVANDILYARSKQSTLLSLVSSKSTFSCRAIPPTWTDFNFPFP

gene

CDS

PERGEHTYQSPVSPENSGSKVRKIQALKKYPSSALLIGRAIGBKWTMKECI
IFGLCAKASTKEECVCHGTRALGDIACLLQCRHPRIVELFVGDYSNLLLEHVE
YGCIRNFMASEAPLDTFETVRICKDIYSAMAYLESIRIVHGHFTPNILMDGFHAK
VCSPPSHHAQRLRYSAPEIVNNEFTHKSDAWAVATVYEMAYQCRORPPEELTNEQ
IVDNACALLDQPNNAVVPPLMPTVFNVEILLTLCFRVDQLERTFERLVPRQD"
22629..27352
/gene="C25F6.3"
join(22629..22778,22996..23109,23156..23244,23886..24135,
24185..24653,24703..24875,24923..25050,25114..25196,
25344..25495,25544..26161,26225..26956,27003..27136,
27190..27352)
/gene="C25F6.3"
/note="coded for by C. elegans cDNA yk28e9.3; coded for by
C. elegans cDNA yk28e9.5; coded for by C. elegans cDNA
yk28h9.3; coded for by C. elegans cDNA yk28h9.5; coded for
by C. elegans cDNA yk40e4.5; coded for by C. elegans cDNA
yk5h3.5; weakly similar to dihydroorotate dehydrogenase
and glutamate synthase"
/codon_start=1
/protein_id="AA080433.1"
/db_xref="GI:1049458"
/translation="MSVDFYAFCLSIFFICKFFNIIFTAMTPKNTTSPNNPLLSK
DSPDISLLINPKVQDKANAVSAVTKKKNHKNRNEKCGCSTGESKLKNFRDI
KHTLTSERGALKAMRCLCADAPQKSCPTQLDVKSFITSISNKNYIGARQILSDN
PLGLTCMIGCPTSLDCNLOAEEGAINIGLQOYACDFKVMNQVITVSKVEVE
NRNASHKEQVALICGCPASIFASFLARLGYTDITYEKRAYIGLSAEIPQRLPY
WDFEIOIARDIGVOIETNRPLGDKGLTAKLKEOGAAVFIGINPEIPDLFEG
LTENGFTYSKNYLPVAAASKPCMGCKRTPLPTMRGVVVLGAGDTMPCATSALR
CGASRTIARFKGTGIRAVPESMEAKEKCELPESAPRKINVKDGRVISEFNT
EQDNKWEDEEQIVILKCDYVISAFGSLKEDAVLSAQPCQLNKNWGLVDSSTQ
QTSKRWKAGVDGVAETTESVNDGKIAMNNHRYIQSLHGNQVSETPELPQFTF
IDEDVLSVDMCGVGFENPFLGASAPPTTSGPMCRRAPEQGWGFTITTYGDLQVLT
VSPRIVRGSTGPLYGPNQGSFMNIELISEKSCYQLQICRELKRDHPTKIVIASIMC
VYNKADWIELATSEAGADILENLSCPHMGEGKGLACGQSPVKEICRWVRAC
KTYFPKPTPNITDVREIARAARDGASGVATNTVYSLMHMKADGNAPWAIQSTKR
TYTGMGSGSAIRPIAMKANSSIANELDGFPIMATGIESATETGLGLFMAGASVQVCS
AVONQDTFVDDYCTGLKALLYLSGAESNLDTNHFASRPDTQVSRVPTVEDVIGKALPQVY
FGYRDEREKEAIKISENLDTNHFASRPDTQVSRVPTVEDVIGKALPQVY
TLNQKQVIAIIDDMCINGCYMTCDNSGYQAITFDPVTHQPHVEDDCTGCTLCY
SVCPIPECIEMVPTGPKAKRGKVPSEVPTPVVVKVQDQGRVILDTTGMQ"
34301..39725
/gene="C25F6.2"
join(34301..34613,35496..35812,35889..36107,36155..36251,
36299..36389,36443..36794,36851..37157,37216..37684,
37743..38266,38350..38632,38680..38801,38857..38870,
38931..38949,39228..39253,39483..39725)
/gene="C25F6.2"
/note="coded for by C. elegans cDNA yk25e5.3; coded for by
C. elegans cDNA yk25e5.5; similar to repeat guanylate
kinase domain of D. melanogaster lethal(1) discs large-1
tumor suppressor protein (Sp:DLGLDROME, P31007) and R.
norvegicus postsynaptic density protein 95 (PSD-95)
(Sp:PSD9_RAT, P31016)"
/codon_start=1
/protein_id="AA080434.1"
/db_xref="GI:1049459"
/translation="MAALCVQVCACVCKRFLFLSDKLCILIAAQFFLFFVNASFLP
SSYVERLEALKPVVYFQSKLVDFOPNFNFORPASKENSEWEERLTREGSSHES
SEKAKHAIENVEDYQCTLRHNEELTNLERTVTFKSNLMHSLDILHDLYEQTLLS
ERKSDAEKNNEVRVIERLEGGSYNSRPAANTYSTNYLSSTTPLLSDLRDGGFS
YLNKGGLNGLNGLLSSPSSSTHYLHERQRTSHDGTWRTTVDTPG
LERRVVDHTGIDDDHGRKWELENIVLEKHTGLGFSITGMQDQEDGKTRDIYVNI
EGGAALADGRMKNDIITAVNTNENCNKHEAVNALKSGSNVSLSLKRRKDEAFLLP
IGNFGGSTYLSRGVTPPSAGNLOHAIHSPSAPIHPPPPPPVHHGSLSOLSGVQYR
STPNTSVIDLKVGARGLGFSLAGGQNEHVKGDTDIYVTKIIEGAELDGLRVLG
KILEVDHSLINTHENAVNLKNTGNRVLIIQOQTGAIFENDASQOFMPTTPILRP
SSVDYNSQSGSHLSYGGPLNTSYGSLVAVNGVLRNATHKEAALRAGNPNVYL
EDNEPTIYSFVPGGVADLSYKNTGDVLLVAVNGVLRNATHKEAALRAGNPNVYL
TLRYRQYQIFESKEKLRNDVIAQRMGTLSRKSYVRFALDYDFSRNSVGNFY
MGNVQDILHIINSDDMTARKVHENGTEAGVIPSKKRKRKRRLKRVQNFNS
GSOIACRNSSTTGLENPRGSRQSOLSEKFPFVKSTDLNDLNESSNAVEPWSYQ
AVSQALNVYRPIIILGALDRINDELVLNRDPSKFSSCVSTDLNDLNESSNAVEPWSYQ
VNKHMEEDVYNLFIAGQFQNNLYGTSIQSVKDVANQGRHCLLDVYSGNARILQSN


```

CDS
Query Match      8.1%; Score 70; DB 4; Length 303092;
Best Local Similarity .63.9%; Pred. No. 0.00027;
Matches 106; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 267 AGGACATTGTGACAAATGTTTCCCCCAATAATCATCCGGGGAACCACTCTCGCCCCCATGT 326
      |||| | || || || || || || || || || || || || || || || || || || ||
Db 165897 AGGATCTGGTCACGAATGCTCGCGCGCATCGTCAGGGGCACCACTCGGGCTACAAGT 165838

Qy 327 ATGGCCCTGGACAAAGCTTCCTTCTGATATTGAGCTCATCAGTGAGAAAACGGCTGCAT 386
      |||| | || || || || || || || || || || || || || || || || || || ||
Db 165837 ATGGACCCGACAGGGCTGTGTTCTCGACATCGAACAATCATCGGGAAGCGGCCGAGT 165778

Qy 387 ATTGGTGTCAAAAGTGTCACTGAAGCTAAAGCGTGACTTCCCAAGACAA 432
      |||| | || || || || || || || || || || || || || || || || || || ||
Db 165777 ACTGGCTGAATCGATTGGAGAAGTGAAGCGTGACTTCCCGGAGAA 165732

RESULT 15
DMU65491      2348 bp mRNA INV 26-SEP-1996
LOCUS      Drosophila melanogaster Dreg-3 protein mRNA, complete cds.
DEFINITION
ACCESSION      U65491
VERSION      U65491.1 GI:1561729
KEYWORDS
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS      Van Gelder, R.N., Bae, H., Palazzolo, M.J. and Krasnow, M.A.
TITLE      Extent and character of circadian gene expression in Drosophila
              melanogaster: identification of twenty oscillating mRNAs in the fly
              head
JOURNAL      Curr. Biol. 5 (12), 1424-1436 (1995)
MEDLINE      96362140
REFERENCE      2 (bases 1 to 2348)
AUTHORS      Van Gelder, R.N., Bae, H., Palazzolo, M.J. and Krasnow, M.A.
TITLE      Direct Submission
JOURNAL      Submitted (29-JUL-1996) Ophthalmology and Visual Sciences,
              Washington University School of Medicine, 660 S. Euclid Avenue, Box
              8096, St. Louis, MO 63110, USA
FEATURES
Source      1..2348
              /organism="Drosophila melanogaster"
              /strain="Canton-S"
              /db_xref="taxon:7227"
              358..2238
              /note="similar to dihydropyrimidine dehydrogenase;
              rhythmically expressed gene 3"
              /codon_start=1
              /product="Dreg-3 protein"
              /protein_id="AAC47288.1"
              /db_xref="GI:1561730"
              /translation="MEFCRTEQNEDEWEDEBQTORLKANFVISAFGSGLEDDQVKA
              ALAPLQFGLPVVDRTMQSSVKQVFLGDLGAVNTTVESVNDCKVAWSIHCQLQ
              GLPDTPTAALPLFTDIDVDISVEMCGIRFENPFGLASAPPTTSTAMIRAEQGWG
              FVYTFGLDKDLVNTVPRVIRGTTSGYKPPQGCFLNIELISEKRAEYWLKISGE
              LKRDFFPKIVIASINCSFNEDWTEIAKAEQSGADALELNLSCPHGMGERGMGLAG
              QDPVELVQISRWKAVKLPFTTKLTPNIDVSIAAAARKREPRMSAINTVOGLMGL
              KADSTAWPAIGKEQRTYGGVSGNATPMALKALSDIANRVGPEPILIGGIDSGEVA
              LQTHAGATVLIQICSSVQNDFTVIEDYCTALKALLYLKRIHQSWVPVSNASHHPRR
              SIRQRACPPDRRGKATLGFPGYQQRORDIKMAELRSQKGLSWDAEQVKATPPASNG
              APNPAPRIKDVIGALDKIGSYNKLKQOKVALIDDDMCINGCKYMTCADSGYQAI
              EFDKDTHTHPVNDCTGCTLCVSVCPIDICTITWPKKIPHVIRKGVEEKIFYTHALSO
              CQ"
BASE COUNT      538 a 649 c 684 g 477 t
ORIGIN
```

Query Match

7.9%; Score 68.4; DB 6; Length 2348;

Search completed: November 20, 2001, 16:01:12
Job time: 5611 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 14:29:36 : Search time 123.96 Seconds
(without alignments)
4361.272 Million cell updates/sec

Title: US-09-308-080-1

Perfect score: 861

Sequence: 1 TGTATTAAGATAAATAATT.....ACTGGGAATAATTATTATAA 861

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	857.8	99.6	861	AAT91798	Dihydropyrimidine
2	171.2	19.9	3951	AA222903	Human DPD DNA. Ho
3	169.6	19.7	3957	AAT14077	Human dihydropyrim
4	169.6	19.7	4368	AAT03133	Human dihydropyrim
5	169.6	19.7	4369	AAT03143	Human dihydropyrim
6	153.4	17.8	4414	AAT03132	Bovine dihydropyri
7	153.4	17.8	4430	AAT03142	Bovine dihydropyri
8	151.4	17.6	936	AAF58252	Oligonucleotide D1
9	151.4	17.6	936	AAF58254	Oligonucleotide D1
10	151.4	17.6	936	AAF58257	Oligonucleotide D1
11	151.4	17.6	936	AAF58259	Oligonucleotide D2

12	151.4	17.6	936	22	AAF58262	Oligonucleotide D2
13	151.4	17.6	938	22	AAF58255	Oligonucleotide D1
14	150.2	17.4	4447	17	AAT14078	Pig dihydropyrimid
15	147.4	17.1	936	22	AAF58252	Oligonucleotide D1
16	147.4	17.1	936	22	AAF58254	Oligonucleotide D1
17	147.4	17.1	936	22	AAF58257	Oligonucleotide D1
18	147.4	17.1	936	22	AAF58259	Oligonucleotide D2
19	147.4	17.1	936	22	AAF58262	Oligonucleotide D2
20	147.4	17.1	938	22	AAF58255	Oligonucleotide D1
21	61.2	7.1	244	22	AAF58238	Oligonucleotide D1
22	59.8	6.9	244	22	AAF58238	Oligonucleotide D1
23	45	5.2	5852	12	AAQ11710	Dictyostelium plas
24	42.4	4.9	3648	22	AAQ11710	Dictyostelium plas
25	42.4	4.9	3648	22	AAQ11710	Dictyostelium plas
26	42.4	4.9	11461	22	AAQ11710	Dictyostelium plas
27	42.4	4.9	12766	22	AAQ11710	Dictyostelium plas
28	42.4	4.9	13737	21	AAQ11710	Dictyostelium plas
29	42.2	4.9	2400	18	AAQ11710	Dictyostelium plas
30	42.2	4.9	8318	20	AAQ11710	Dictyostelium plas
31	41.4	4.8	745	20	AAQ11710	Dictyostelium plas
32	41	4.8	3947	19	AAQ11710	Dictyostelium plas
33	41	4.8	3947	19	AAQ11710	Dictyostelium plas
34	40.2	4.7	2313	21	AAQ11710	Dictyostelium plas
35	40	4.6	37808	20	AAQ11710	Dictyostelium plas
36	39.4	4.6	1717	20	AAQ11710	Dictyostelium plas
37	39	4.5	14752	20	AAQ11710	Dictyostelium plas
38	39	4.5	116277	20	AAQ11710	Dictyostelium plas
39	39	4.5	910715	20	AAQ11710	Dictyostelium plas
40	38.8	4.5	5810	18	AAQ11710	Dictyostelium plas
41	38.8	4.5	8911	20	AAQ11710	Dictyostelium plas
42	38.6	4.5	1400	20	AAQ11710	Dictyostelium plas
43	38.6	4.5	1591	21	AAQ11710	Dictyostelium plas
44	38.6	4.5	111309	20	AAQ11710	Dictyostelium plas
45	38.4	4.5	1776	18	AAQ11710	Dictyostelium plas

ALIGNMENTS

RESULT 1

AAT91798
ID AAT91798 standard; DNA; 861 BP.

XX

AC AAT91798;

XX

DT 14-APR-1998 (first entry)

XX Dihydropyrimidine dehydrogenase genomic fragment.

DE Human; dihydropyrimidine dehydrogenase; DPD; slicing defect;

KW detection; 5-fluorouracil; cancer; anticancer; uraciluria; ds.

KW

XX Homo sapiens.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

This sequence represents the DNA sequence encoding human dihydropyrimidine dehydrogenase (DPD). DPD catalyses the initial and rate limiting step in pyrimidine catabolism, the reduction of pyrimidines to 5,6-dihydropyrimidines. DPD is a complex enzyme consisting of two identical subunits, containing FMN, FAD and iron-sulphur centers, and utilising NADPH as a cofactor. DPD has also been shown to catalyse the reduction of various pyrimidine analogues including the fluoropyrimidine anticancer drug 5-fluorouracil (Fura). Up to 85% of administered Fura may be catabolised by DPD, and it therefore governs the effectiveness of Fura as an anticancer drug. DPD genes or fragments of them may be used in the detection of DPD in a sample, esp. isolated from a cancer patient. According to the amount of DPD detected, a therapeutically effective amount of Fura may be determined and administered. DPD deficiency, leading to life-threatening toxicity on exposure to Fura, in a human caused by a frameshift mutation may be determined by means of a molecular biological assay to detect the deletion of an A residue at codon 318 within the DPD-coding region.

SQ Sequence 4368 BP; 1325 A; 829 C; 932 G; 1282 T; 0 other;

Query Match	19.7%	Score	169.6	DB	16	Length	4368
Best Local Similarity	92.7%	Pred. No.	5.5e-29				
Matches	178	Conservative	0	Mismatches	14	Indels	0
Gaps	0						
QY	244	CTCAATATCTTTTACTCTTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAAATCATCG	303				
Db	1764	caccaaaacttctctcttgataaggacattgacaaatggtttcccccagataatcgcg	1823				
QY	304	GGSAACCACTCTGCGCCCATGTATGCCCCGGAACAAAGTCCTTTCTGAATATTGAGCT	363				
Db	1824	gggaacacactctggcccatgatatggccctggacaaagctcttctgaaatatgagct	1883				
QY	364	CATCATGTGAGAAACGGCTGCATATTCGGTGCAAAAGTGTCACCTGAACATAAGCGCTGACTT	423				
Db	1884	catctcgagaaaacggtgcataatctggtgccaaagtgcactgaaactaaaggctgacct	1943				
QY	424	CCAGACAACGT	435				
Db	1944	tccagacaacat	1955				

RESULT	5	
AAT03143		
ID	AAT03143	standard; cDNA; 4369 BP.
XX		
XX	AAT03143;	
XX		
XX	05-JUN-1996	(first entry)
XX		
XX	Human dihydropyrimidine dehydrogenase cDNA.	
DE		
DE		
KW	Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;	
KW	pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;	
KW	fluoropyrimidine; anticancer drug; 5-fluorouracil; Fura; cancer;	
XX	frameshift mutation; ss.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FH	49..3125	
CDS		
FT	/*tag= a	
FT	/product= Human lymphocyte DPD	
FT	complement (36..60)	
FT	/*tag= b	
FT	primer_bind	
FT	primer_bind	
FT	complement (716..743)	
FT	/*tag= c	
FT	primer_bind	
FT	1260..1288	
FT	/*tag= d	
FT	3117..3139	
FT	primer_bind	
FT	/*tag= e	
FT	mutation	
FT	1000	

FT		/tag= f	
FT		/note=	"Position of single base deletion in DPD
FT			deficient patients"
PX			
PN	WO9528489-A1.		
XX			
PD	26-OCT-1995.		
XX			
PF	13-APR-1995;	95WO-US04567 .	
XX			
PR	13-APR-1994;	94US-0227357 .	
XX			
PA	(UABR-) UAB RES FOUND.		
PI	Cheng X, Diasio RB, Johnson M,	Lu Z, Zhang R;	
XX			
DR	WPI; 1995-373803/48.		
DR	P-PSDB; AAR86001.		
XX			
PT	Novel dihydro:pyrimidine dehydrogenase gene - used to optimise		
PT	5-fluoro:uracil doses given to cancer patients		
XX			
PS	Disclosure; Fig 7; 207pp; English.		
XX			
CC	This sequence represents the cDNA sequence encoding human dihydro-		
CC	pyrimidine dehydrogenase (DPD). This sequence was amplified using		
CC	the primer sequences given in AA03144-47. DPD catalyses the initial		
CC	and rate limiting step in pyrimidine catabolism, the reduction of		
CC	pyrimidines to 5,6-dihydropyrimidines. DPD is a complex enzyme		
CC	consisting of two identical subunits, containing FMN, FAD and iron-		
CC	sulphur centers, and utilising NADPH as a cofactor. DPD has also been		
CC	shown to catalyse the reduction of various pyrimidine analogues		
CC	including the fluoropyrimidine anticancer drug 5-fluorouracil (Fura).		
CC	Up to 85% of administered Fura may be catabolised by DPD, and it		
CC	therefore governs the effectiveness of Fura as an anticancer drug.		
CC	DPD genes or fragments of them may be used in the detection of DPD in a		
CC	sample, esp. isolated from a cancer patient. According to the amount		
CC	of DPD detected, a therapeutically effective amount of Fura may be		
CC	determined and administered. DPD deficiency, leading to life-		
CC	threatening toxicity on exposure to Fura, in a human caused by a		
CC	frameshift mutation may be determined by means of a molecular biological		
CC	assay to detect the deletion of an A residue at codon 318 within the		
CC	DPD-coding region.		
XX			
SQ	Sequence 4369 BP; 1325 A; 829 C; 932 G; 1283 T; 0 other;		

Query Match	19.7%;	Score 169..6;	DB 16;	Length 4369;
Best Local Similarity	92.7%;	Pred. No. 5.5e-29;		
Matches 178;	Conservative	0;	Mismatches 14;	Indels 0; Gaps 0;
QY 244	CTCAATATCTTTACTCTTTTCATGTAGGAGCATGTGCACAAATGTTTCCCCCAATAATCATCCG	303		
Db 1764	caccaaaacttctctcttgataaaggacattgtgacaaaattttcccccaagaacatccg	1823		
QY 304	GGGAACCACTCTCTGGGCCCATGTATNGCCCTGGACAAAGCTCTTTCGTGAATATTGAGCT	363		
Db 1824	gggaaccacactcttgccccatgtatggccctggacaaagctccctttctgataattgagct	1883		
QY 364	CATCAGTGTAGAAAACGGCTGCATATTGGTGTCAAGTGTCTACTGAACCTAAAAGGCTGACTT	423		
Db 1884	catcagtgtgaaaaacggctgcataattgggtgccaaagtgcactgaactaaaggctgacct	1943		
QY 424	CCCAGACAACGT	435		
Db 1944	tccaagacaacat	1955		

RESULT 6
AAT03132
ID AAT03132 standard; DNA; 4414 BP.
XX
AC AAT03132;

[illegible]

CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.

SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 17.6%; Score 151.4; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 4.6e-25;
Matches 8; Conservative 511; Mismatches 272; Indels 0; Gaps 0;

QY 9 AAGATAAATATTTTTCGTTCTTAAACCTAGGTTTACAAAGTAATTTATC 68
DB 2 WWW 61
QY 69 TGGAGCTAACAAATCTTTTACCTTTTATTTGCAAGTAGTTTATGTTCAATCTA 128
DB 62 WWW 121
QY 129 ATTTAATGTATATAAATCTCTGCAATATGTGAGGAGGACCTCATAAATATTG 188
DB 122 WWW 181
QY 189 TCATATGGAATGAGCAGATAATAAAGATTATAGCTTTTGTCAAAAGGAGACTCA 248
DB 182 WWW 241
QY 249 TATCTTTACTCTTTTCATGAGGACATTTGACAAATGTTTCCCCCATATATCCTCGGAA 308
DB 242 WWW 301
QY 309 CCACCTCTGCCCCCATGTATGCCCTGGACAAAGCTCCTTCTGATATTTGAGCTCATCA 368
DB 302 WWW 361
QY 369 GTGAGAAAACGGCTGCATATTTGGTGTCAAAAGTGTCACTGAAGCTAAAGGCTGACTTCCAG 428
DB 362 WWW 421
QY 429 ACAACGTAAGTGTGATAAATACTAAACAAGAGATTGGCATAAGTTGGTGAATGTTTA 488
DB 422 WWW 481
QY 489 TTAAACATCCAATTCATAGGCTTTAATAATATAATGTTATTTATCAACGAATCTG 548
DB 482 WWW 541
QY 549 CCAGTTGCTTTGCTGATGATAGAAGATAAAGAAAGAAAGAAAGCTCAAGAACTATA 608
DB 542 WWW 601
QY 609 AAACCCACACAATGTGAAGCTCTGTATAAATGGTGCCATGTAAGATGGAAGAGTATC 668
DB 602 WWW 661
QY 669 TACATAGCAGAGGAAGAGAAATGAATCTCAATTTTATTTGAGTTGGCCCCCAGTAT 728
DB 662 WWW 721
QY 729 GTGGCTGGTATTTATCAGGCTGATGCCCGCAGGAGAAATTTGAACACTATAACCACTCCA 788
DB 722 WWW 781
QY 789 AATATAAACC 799
DB 782 WWW 792

RESULT 11
AAF58259
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX

24-APR-2001 (first entry)
Oligonucleotide D2004.
Electron-transfer group; ETM; mismatch; genotyping;
gene expression; ss.
Synthetic.
WO200107665-A2.
01-FEB-2001.
26-JUL-2000; 2000WO-US20476.
26-JUL-1999; 99US-0145695.
17-MAR-2000; 2000US-0190259.
(CLIN-) CLINICAL MICRO SENSORS INC.
Umek RM;
WPI; 2001-159728/16.
Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -
Example 6; Page 128; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 17.6%; Score 151.4; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 4.6e-25;
Matches 8; Conservative 511; Mismatches 272; Indels 0; Gaps 0;

QY 9 AAGATAAATATTTTTCGTTCTTAAACCTAGGTTTACAAAGTAATTTATC 68
DB 2 WWW 61
QY 69 TGGAGCTAACAAATCTTTTATTTACCTTTTATTTGCAAGTAGTTTATGTTCAATCTA 128
DB 62 WWW 121
QY 129 ATTTAATGTATATAAATCTCTGCAATATGTGAGGAGGACCTCATAAATATTG 188
DB 122 WWW 181
QY 189 TCATATGGAATGAGCAGATAATAAAGATTATAGCTTTTGTCAAAAGGAGACTCAA 248
DB 182 WWW 241
QY 249 TATCTTTACTCTTTTCATGAGGACATTTGACAAATGTTTCCCCCATATATCCTCGGAA 308
DB 242 WWW 301
QY 309 CCACCTCTGCCCCCATGTATGCCCTGGACAAAGCTCCTTCTGATATTTGAGCTCATCA 368
DB 302 WWW 361
QY 369 GTGAGAAAACGGCTGCATATTTGGTGTCAAAAGTGTCACTGAAGCTAAAGGCTGACTTCCAG 428
DB 362 WWW 421
QY 429 ACAACGTAAGTGTGATAAATACTTAAACAAGAGAAATTTGGTGAATGTTTA 488

	Query Match	17.6%; Score 151.4; DB 22; Length 936;
	Best Local Similarity	1.08; pred. No. 4.6e-25;
	Matches 8; Conservative 511; Mismatches 272; Indels 0; Gaps	
QY	9 AAGATAAATATTTTGTTTTTCCTAAACCTAGGGTTACAGAAGTAGTAATTATC 68	
DB	2 WWWWWWWW	61
QY	69 TGGAGCTAACAAATACTTATTTTAACTTTTATTGCAAGTAGTTTATGTTCAAATCTA 128	
DB	62 WWWWWWWW	121
QY	129 ATTATAGTATATAAAAAATCCCTCGCAATATGTGAGGAGGGACCTCATAAAATATTG 198	
DB	122 WWWWWWWW	181
QY	189 TCATATGGAATGAGCAGATAATAAGATTATAGCTTTTCTTGTCAAAAGGAGACTCAA 248	
DB	182 WWWWWWWW	241
QY	249 TATCTTTTACTCTTCATGAGGACATTTGTGACAATGTTTCCCOCATAATCATCGGGGAA 308	
DB	242 WWWWWWWW	301
QY	309 CCACCCTGCCCCCATGTATGCCCTGGCACAAAGCTCCTTTCGTAATTTGAGCTCATCA 368	
DB	302 WWWWWWWW	361
QY	369 GTGAGAAACGGCTGCATATTTGGTGTCAAAGTGTCACTGAACATAAGGCTGACTCCAG 428	
DB	362 WWWWWWWW	421
QY	429 ACAACGTAAGTGTGATAAAATCTAAACAGAGAATTGGCATAAGTTGGTGAAATGTTTA 488	
DB	422 WWWWWWWW	481
QY	489 TTTAACATCCAATTCATAGSCTTATAAATATTAATGCTGTATATTTTATCAACGAACTCG 548	
DB	482 WWWWWWWW	541
QY	549 CCAGTTGCTTTGCTGATGCATAGAAAGATAAAAAGAGAAAGCTCAAGAACTCATAA 608	
DB	542 WWWWWWWW	601
QY	609 AAACCCACACAATGTGAAGCTCTGTTATAAATGGTGCCCATGTAAGATGGAAGAAGTATC 668	
DB	602 WWWWWWWW	661
QY	669 TACATAAGCAGAGGAAGAGAAATGAATACTCATTTTATTGAGTTGGCCCCCAGCTGAT 728	
DB	662 WWWWWWWW	721
QY	729 GTGGCTGGTATTTATGAAGTGTGATGCCCCAGGAAGAAATTTGTAACCTATAAACCTCCA 788	
DB	722 WWWWWWWW	781
QY	789 AATATAAACCC 799	
DB	782 WWWWGCCCC 792	

22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 5

Qy 369 GTGAGAAAACGGCTGCATATTGGTGTCAAAGTGTCACCTGAAGCTAAAGGCTGACTTCCCAG 428

[illegible]

Db 362 www.wwwwwwwwwwwwwgcta wwwwwwwwwwwwwww 421

100
90
80
70
60
50
40
30
20
10
0

QY 429 ACAACGTAAGTGTGATAAAAATCTAAAAACAAGAGAAATGGCATTAAGTGGTGAAATTTA 488

Db 422 481

QY 489 TTTAAACATCCAATTTCATAGGCTTATAAAATATTAAATGTTATATTTTATCAACGAATCTG 548

pk 483 541

DD 48Z WWWWWW C WWWWWW WWWWWW 341

QV 549 CCAGTTGCTTTGCTGATGATAGAAAGATAAAAAAGAAAGCTCAAGAACTCATAA 608

[illegible][illegible]

Q	C
6	E
6	E
	A
	A
	C
	E
	A
	A
	E
	E
	C
	C
	E
	C
	E
	A
	A
	E
	E
	E
	C
	C
	E
	E
	A
	C
	C
	A
	A
	C
	C
	C
	A
	A
	C
	C
	: :

QY 609 AAACCCACACAAATGTTGAAGCTCTCTGTTATATAAATGGGTGGCCATGTAAAGATGGAAGAAATC 609

Db 602 www..... 661

.....

QY 669 TACATAAGCAGAAGGAGAAATGAAATACTCATTTTATTGAGTTGCCCCCACTGTAT 728

[illegible]

DB 602 WWWWWW C WWWWWW / ZT

QV 729 GTGGCTGGTATTATGAAGGTGATGACCCAGGAAGAAATTGTAACCTATAAACCACTCCA 788

[illegible][illegible]

0
0
7
C
C
A
A
E
A
A
C
7
:

QY	789	AAATATAAACC	799
	 III	
Db	782	wwwggccc	792

RESULT 13

AAF58255

ID AAF58255 standard; DNA; 938 BP.

XX
XX
7A058255.

AC
HAF 50255,
XX

DT 24-APR-2001 (first entry)

XX

DE Oligonucleotide D1876.

Electron-transfer group: ETM: mismatch: genotyping:

EW
ELECTION TRANSFER GROUP, ERM, MISMANAGE, GENOCYDING,

CC 4-fluorouracil.

Sequence 4447 BP; 1285 A; 928 C; 1000 G; 1234 T; 0 other;

Query Match	Score	DB	Length
17.48	150.2	17	4447

Best Local Similarity 87.7%; Pred. No. 1.2e-24;

Matches	164;	Conservative	0;	Mismatches	23;	Indels	0;	Gaps	0;
----------------	-------------	---------------------	-----------	-------------------	------------	---------------	-----------	-------------	-----------

QY 246 CAATATCTTTACTCTTTCATGAGGACATTGTGACAAATGTTTCCCCCATAATCATCCGGG 305

7

Db 1805 ccaaaacttctcttgataaggacatagtgacaaatgtctcaccagaatcgctccggg 1864

.....

QY 306 GAACCACTCTGGCCCCATGTATGGCCCTGGACAAGCTCCTTTCTGAATATTGAGCTCA 365

Db 1865 ggactacctgtgcccccatgtacggccctggacaaagctccttctcctgaatattgagctca 1924

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

RESULT 15

AA58252/C

ID AAF58252 standard; DNA; 936 BP.

[illegible]

AC AAF58252;

[illegible]

DT 24-APR-2001 (first entry)

XX
XX

DE Oligonucleotide D1835.

XX-

KW Electron-transfer group; ETM; m

KW gene expression; ss.

XXC

OS Synthetic.

XX
DN
13030010766-30

PN WO200107665-A2.

XX
DD 01-FEB-2001

FD
XX
01-FEB-2001.

XX
PF
26-JUL-2000: 2000WO-VIS20476

PF
XX
26-JUL-2000; 2000WW0-US20476.

XX 26-JUL-1999: 99US-0145695.

20-JUL-1999; 990US-0143693.
17-MAR-2000: 2000US-0190259.

XX
XX
17 - MAR - 2000, 2000S - 0150239.

PA (CLIN-) CLINICAL MICRO SENSORS

XX
 (CONTINUED)
 MICROSENSORS

PI Umek RM;

XX
XX
XX

DR WPI; 2001-159728/16.

XX
XX

Query Match

17.18: Score 147.4: PB 22: Length 936:

Search completed: November 20, 2001, 15:59:11
Job time: 5375 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2001, 14:25:51 ; Search time 1355.35 Seconds
(without alignments)
6005.022 Million cell updates/sec

Title: US-09-308-080-1
Perfect score: 861
Sequence: 1 TGTTAATGAAGATAAATATT.....ACTGGGAATAATTATTATAA 861

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	169.6	19.7	938	172	BG024298	602274460
2	92.4	10.7	853	220	CNS02EEI	Tetraodon
3	86.2	10.0	621	110	AW018806	fd6d0d3.y
4	70	8.1	1207	219	CNS00175	Drosophil
5	62.6	7.3	562	165	BE288474	601095732
6	58	6.7	1101	219	CNS0039G	Drosophil
7	55	6.4	500	107	AU086287	AU086287
8	53.4	6.2	483	151	BF653737	277669 MA
9	52.6	6.1	1101	219	CNS00EVL	Drosophil
10	52.6	6.1	1101	219	CNS0020C	Drosophil
11	52.4	6.1	1025	219	CNS014J2	Drosophil
12	52	6.0	912	245	A2527678	ENTEN23TR
13	52	6.0	1101	219	CNS00EO7	Drosophil
14	51.8	6.0	853	245	A2548198	ENTFW05TF
15	51.8	6.0	868	248	A2686179	ENTJ141TR
16	51.8	6.0	881	247	A2681140	ENTWC34TR
17	51.8	6.0	899	245	A2542627	ENTCX21TR
18	51.4	6.0	898	248	A2686287	ENTRG36TR
19	51.2	5.9	987	219	CNS014PQ	Drosophil
20	50	5.8	632	107	AU060843	AU060843
21	49.8	5.8	947	247	A2682793	ENTKY77TF
22	49.8	5.8	1101	219	CNS016LI	Drosophil
23	49.6	5.8	867	219	CNS00CX5	Drosophil
24	49.4	5.7	872	245	A2535541	ENTBP32TR
25	49.4	5.7	1101	219	CNS0039G	Drosophil
26	49	5.7	896	245	A2534072	ENTBR33TR
27	48.6	5.6	872	245	A2527849	ENTCF51TR
28	48.6	5.6	1101	219	CNS017V6	Drosophil
29	48.4	5.6	329	105	AL513719	AL513719
30	48.4	5.6	911	248	A2686515	ENTID42TR
31	48.2	5.6	898	247	A2674122	ENTIB47TF
32	48.2	5.6	987	219	CNS014PQ	Drosophil
33	48.2	5.6	1007	219	CNS01460	Drosophil
34	48	5.6	892	245	A2549872	ENTFG30TR
35	47.8	5.6	1101	219	CNS003DQ	Drosophil
36	47.4	5.5	872	248	A2683520	ENTIO43TR
37	47.4	5.5	878	219	CNS018R7	Drosophil
38	47.2	5.5	816	220	CNS02KJV	Drosophil
39	47.2	5.5	909	248	A2690408	ENTLW08TR
40	47.2	5.5	1032	220	CNS02OIP	Drosophil
41	47.2	5.5	1048	248	A2683764	ENTIL96TR
42	46.8	5.4	959	219	CNS00655	Drosophil
43	46.8	5.4	1021	219	CNS014DY	Drosophil
44	46.8	5.4	1092	220	CNS020K7	Tetraodon
45	46.8	5.4	1201	219	AL106495	Drosophil

ALIGNMENTS

Result No.	Score	Query Match %	Length	DB	ID	Description
1	602274450F1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
LOCUS	BG024298	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
DEFINITION	BG024298	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
ACCESSION	BG024298	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
VERSION	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
KEYWORDS	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
SOURCE	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
ORGANISM	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
REFERENCE	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
AUTHORS	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
TITLE	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
JOURNAL	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'
COMMENT	BG024298.1	938 bp	NTN_MGC_85	EST	24-JAN-2001	4362302 5'

Email: cgapbs-i@mail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLAM10006 row: h column: 15
High quality sequence stop: 714.
Location/Qualifiers
1. 938

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4362302"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 222 a 234 c 253 g
ORIGIN

Query Match 19.7%; Score 169.6; DB 172; Length 938;
Best Local Similarity 92.7%; Pred. No. 7.9e-29;
Matches 178; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 244 CTCAATATCTTACTCTTCATGAGGACATGTGACAAATGTTCCCCCATATCATCCG 303

Db 260 CACCAAACTTCTCTTGTATAGGACATGTGACAAATGTTCCCCCATATCATCCG 319

QY 304 GGAACACCTCTGGCCCCATGATGGCCCTGGACAAAGCTCCTTCTGTAATATTGAGCT 363

Db 320 GGAACACCTCTGGCCCCATGATGGCCCTGGACAAAGCTCCTTCTGTAATATTGAGCT 379

QY 364 CATCAGTGAGAAACGGCTGCATATTTGGTCTCAAGTGTCTACGTAAGCTGACTT 423

Db 380 CATCAGTGAGAAACGGCTGCATATTTGGTCTCAAGTGTCTACGTAAGCTGACTT 439

QY 424 CCCAGACAAAGCT 435

Db 440 TCCAGACAAAT 451

RESULT 2
CNS02EEI/c

LOCUS

DEFINITION

CNS02EEI 853 bp DNA GSS 13-MAY-2000

Tetraodon nigroviridis genome survey sequence T7 end of clone

262A08 of library G from Tetraodon nigroviridis, genomic survey

sequence.

AL193635

AL193635.1 GI:7831741

GSS: genome survey sequence.

SOURCE

Tetraodon nigroviridis.

ORGANISM

Tetraodon nigroviridis.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.

1 (bases 1 to 853)

Roest-Crolius.H., Jaillon.O., Dasilva.C., Fizames.C., Fisher.C.,

Bonneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and

Weissenbach.J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished

2 (bases 1 to 853)

Roest-Crolius.H., Jaillon.O., Dasilva.C., Bonneau.L., Fisher.C.,

Bernot.A., Fizames.C., Wincker.P., Brottier.P., Quetier.F.,

TITLE Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 853)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.

FEATURES

source
 Location/Qualifiers

1..853
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="262A08"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG262BA04LP1-end : T7"
 BASE COUNT 252 a 189 c 136 g 269 t 7 others
 ORIGIN

Query Match 10.78; Score 92.4; DB 220; Length 853;
 Best Local Similarity 68.5%; Pred. No. 3.8e-11;
 Matches 126; Conservative 1; Mismatches 57; Indels 0; Gaps 0;
 QY 258 TCTTTCATGAGCATTGTGACAAATGTTCCCCCATATCATCCGGGGAACCACTCTG 317
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 409 TGTTCCTCAGACATGTGTAACAATGTCTCTCTGATTTGCGAGGACCACTTCG 350
 QY 318 GCCCATGATGGCCCTGGACAAAGCTCTTCTGTAATTTAGCTCATCAGTCAGAAAA 377
 Db || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 349 GCAATGTGATGGCCAGGTGAGGGTCAATCTCTCAACATGAGCTTTATAGCGAGA 290
 QY 378 CGGCTGCATATGGTGTCAAGTGTCACTGAACCTAAGGCTGATTTCCCGACAGCACTAA 437
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 289 CGGCAGCTTATTTGGTGTCAATCACTAGTCGCGMACTTAAGAGGAGCTTTCCGAATAATGTA 230
 QY 438 GTGT 441
 Db || ||
 229 GTTT 226

RESULT 3

AW018806 621 bp mRNA EST 13-SEP-1999
 LOCUS fd50d03.y1.zbrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
 DEFINITION SW:DPYD_HUMAN Q12882 DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+]
 PRECURSOR ; mRNA sequence.
 AW018806
 AW018806.1 GI:5872336
 EST.
 zbrafish.
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Rasbora; Danio.
 1 (bases 1 to 621)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
 S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.
 WashU Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@watson.wustl.edu

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT

cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
 www.rzpd.de)

Seq primer: T3 ET from Amersham
 High quality sequence stop: 491.

FEATURES

source
 Location/Qualifiers

1..621
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish WashU MPIMG EST"
 /sex="mixed"
 /tissue_type="26 somite embryos, adult livers, shield
 stage embryos"
 /lab_host="XLI-blue MRF"
 /note="Vector: pSPORT1; Site.1: NotI; Site.2: SalI; 1st
 strand cDNA was primed with a Not I - oligo(dT)15 primer
 [5'PGACTAGTTCATGATCGGAGCGCGCCCTTTTCTTTTCTTTT3];
 double-stranded cDNA was ligated to Sal I adaptors (BRL),
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late somitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

BASE COUNT 151 a 169 c 160 g 141 t
 ORIGIN

Query Match 10.0%; Score 86.2; DB 110; Length 621;
 Best Local Similarity 70.6%; Pred. No. 1e-09;
 Matches 115; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 275 GTCAAAATGTTTCCCCCATATCATCCGGGGAACCACTCTGCCCCCATGATGGCCCT 334
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1 GTACCAACGTGTCAACCGAATCGTGGGACAACTCAGTCACATATTCGGTCCT 60

QY 335 GGACAAAGCTCCCTTCTGTAATATTGAGCTCATCAGTGAGAAACGGCTGCATATTGGTGT 394
 Db || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 61 GGTGAGGCTCATTCCTCAACATCGAGCTCATTAGTGAATAAACTGCCGCTTACTGGTGC 120

QY 395 CAAAGTGTCACTGAACCTAAAGGCTGATTTCCCGACAAACGTAA 437
 Db || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 121 AAGAGCGTGGCTGAACCTCAAAAGCCGACTTCCCAAAATATTA 163

RESULT 4

CNS00175 1207 bp DNA GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR03G12 of RPI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION AL058798
 VERSION AL058798.1 GI:4939103
 KEYWORDS fruit fly.
 SOURCE Drosophila melanogaster

ORGANISM
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1207)

AUTHORS
TITLE Genoscope.
JOURNAL Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammoss in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 Location/Qualifiers
 1. .1207
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR03G12"
 /note="end : TET3"
 BASE COUNT 250 a 298 c 328 g 233 t 98 others
 ORIGIN

Query Match 8.1%; Score 70; DB 219; Length 1207;
 Best Local Similarity 63.9%; Pred. No. 5.3e-06;
 Matches 106; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 267 AGGACATGTGCAAAATGTTCCCCCATATATCCGGGGAACACCTCTGTGGCCCATGT 326
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 399 AGGATGTGTGTCAGAAATGTCGCCGCGCATGTCAGGGGCACACGTCGGGCTACAAGT 458
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 327 ATGGCCCTGGACAAGCTCCTTCTCAATATTGAGTCATCATGTGAGAAAACGGCTGCAT 386
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 459 ATGGCCGACGACGAGGCTGTCTCTCAATCAATCAAGTAATCTCGGAGAGCGGCGGAGT 518
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 387 ATGTGCTCAAACTGTCACCTGAACAAAGCTGCACCTCCAGACAA 432
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 519 ACTGGCTGAATCGATTGGAGAACTGAAGCTGACTTCCCGGAGAA 564
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 5
BE288474
LOCUS 562 bp mRNA EST 26-OCT-2000
DEFINITION 601095732F1 NCL_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3490423 5',
 mRNA sequence.
ACCESSION BE288474
VERSION BE288474.1 GI:9168194
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 562)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM8532 row: p column: 08
 High quality sequence stop: 562.

FEATURES
 Location/Qualifiers
 1. .562
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone="IMAGE:3490423"
 /clone_lib="NCL_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
 Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 137 a 124 c 166 g 135 t
 ORIGIN

Query Match 7.3%; Score 62.6; DB 165; Length 562;
 Best Local Similarity 94.2%; Pred. No. 0.00025;
 Matches 65; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 364 CATCACTGAGAAAACGGCTGCATATTGGTGTCAAGTGTCACTGAACCTAAAGGCTGACTT 423
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1 CATCACTGAGAAAACAGCTGCATATTGGTGTCAAGTGTCACTGAACCTAAAGGCTGACTT 60
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 424 CCGAGACAA 432
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 61 CCGGACAA 69
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 6
CNS0039G
LOCUS 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
 fly); genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
JOURNAL Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammoss in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 Location/Qualifiers
 1. .1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

```
/clone_lib="RPCI-98"
/clone="BACR08K10"
/notice="end : TET3"
BASE COUNT      201 a      64 c      131 g      202 t      503 others
ORIGIN

Query Match      6.7%; Score 58; DB 219; Length 1101;
Best Local Similarity 20.4%; Pred. No. 0.003;
Matches 116; Conservative 216; Mismatches 232; Indels 6; Gaps 1;

QY 70 GGAGCTACAAATACCTTTATTTTACCTTTTATTTGCAAGTAGTTTATGTTCAATCTAA 129
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 391 GGGGATATAWMMWMTTWTTTTANAANAANAATAATTTWAAWAAAAAATTTWAAA 450
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 130 TTTATGTATATAAAATCTCTGCAATATCTGAGGAGGACCTCATAAAAATTTGT 189
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 451 AWAANAATWATWTTTAAWAAAAAANAATTTTWTTTTAAWTTWATWATWTTW 510
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 190 CATATGGAATCAGCAGATAATAAGATTATAGCTTTTCTTCTCAAAAGGAGACTCAAT 249
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 511 WTTAAWAAAAAANAANAANAATAATTTWAAWAAAAAATTTWAAA 570
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 250 ATCTTTACTCTTCATGAGGACATTCGACAAATGTTTCCGCCATATCATCGGGGAAC 309
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 571 YHTTYYTYHYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTY 630
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 310 CACCTCTGCCCTCATGATGCCCTGCACAAAGCTCTCTTCTGAATATTCAGCTCATCAG 369
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 631 HWTYYTAAAYYYTCMYYYHYHMHNAHNAHNAWTTTHTWTHYHWTYHYHYMYCAM 690
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 370 TGAGAAACGGCTGCATATTCGTCTCAAGTGTCACTGAACACTAAAGCTGACTTCCAGA 429
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 691 MCMCTHTCHCYVYHYHYHTAHTHTHHWYAHYMYWYMYWYMYWYMYWYMYWYMYWY 744
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 430 CAACCTAGTGTGATAAAATCTAAACAGAGAAATTTGGCATAAGTTGGTCAATGTTAT 489
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 745 HHYHYHWAHTTWYAWAHAMWMMHHYHAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 804
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 490 TTAACATCCCAATCATAGGCTTAAATATTAATGTGTATATTTATCAACGAATCTGC 549
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 805 YTCYMYTHCHYHYHYHTAYTCWTWTHWMMWTTWTHWMMWTTWTHWMMWTTWTHWMMW 864
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 550 CAGTTGCTTCTGCATGATAGAAAGATAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 609
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 865 WHATTWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHW 924
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 610 AACCCACAAATGTGAAGCTCTGTATAAA 639
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 925 MTCWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 954
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
AU086287 500 bp mRNA EST 27-JAN-2001
LOCUS AU086287 Sugano Malaria cDNA library Plasmodium falciparum cDNA
DEFINITION clone xPfn2249, mRNA sequence.
ACCESSION AU086287
VERSION AU086287.1 GI:12388428
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE 20574754
COMMENT Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@manage.ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES
Location/Qualifiers
source 1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone="xPfn2249"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 214 a 25 c 24 g 231 t 6 others
ORIGIN

Query Match 6.4%; Score 55; DB 107; Length 500;
Best Local Similarity 49.3%; Pred. No. 0.014;
Matches 136; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 7 TGAAGATAATATTTTGTGTTTTCGCTGTTCTAAACCTAGGGTTACAAGAGTAATTTA 66
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 TTAATATATTTTTCATTTTGTAAATATAATATATATTTGACTACATTTATTTAT 261
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 TCTGGAGCTAAACAAATCTTTTATTTACCTTTTATTTGCAAGTAGTTTATGTTCAATTC 126
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 262 TATATTTTCACATATAATTTTATTTTATTTTAAAGAATTTNTATTAAGTTCAAAAT 321
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 127 TAATTAATGTATATAAAATTCCTCTGCAATATGTGAGGAGGACCTCATAAATAT 186
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 322 AAATAAAATCCATATATATATATAATAATAAAACATTTTATAATAATATATAAAAT 381
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 187 TGTCAATGCAATGACGAGATATAAGATTATAGCTTTTCTTTGTCAAAGAGACTC 246
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 382 CTCCTTATAACAAATACAAATAATATATTTATATAATATATAGATATATTTTGTGC 441
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 247 AATATCTTACTCTTCATGAGGACATTTGTGACAA 282
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 442 ATTTTNTGTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 477
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
BF653737 483 bp mRNA EST 20-DEC-2000
LOCUS BF653737 3BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BF653737
ACCESSION BF653737
VERSION BF653737.1 GI:11918869
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fabrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
,W.W. and Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904 e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
```

BACKWARD: GTTTCACGTCACGACG
Plate: 70 row: B column: 4
Seq primer: ATTAGTGACACTATAG.

FEATURES

Location/Qualifiers
1..483
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10p"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

BASE COUNT 128 a 105 c 138 g 112 t

ORIGIN

Query Match 6.2%; Score 53.4; DB 151; Length 483;
Best Local Similarity 90.7%; Pred. No. 0.032;
Matches 68; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 359 GAGCTCATCATGTC-AGAAACGGCTGCATATTGGTGTCAAGTCTCACTGAACCTAAAGGC 417

Db 1 GCGCTCATCATGTGAAAAAACGGCTGCGTATTTGGTGTCAAGTCTCACTGAACCTAAAGGC 60

QY 418 TGACTTCCAGACAA 432

Db 61 CGACTTCCAGACAA 75

RESULT 9

CNS000EVL/c

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1..1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR29B23"

/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 6.1%; Score 52.6; DB 219; Length 1101;
Best Local Similarity 35.7%; Pred. No. 0.051;
Matches 99; Conservative 51; Mismatches 127; Indels 0; Gaps 0;

QY 1 TGTAAATGAAGATAAATATTTTTCGCTGGTCTTAAACCTAGGTTACAGAAGT 60

Db 723 TATWATATWATWATWATTTTTTTTTTTTTTTTAAATATTTAAATWAAAT 664

QY 61 AATTATCTGAGCTAACAAATACTTTTATTTACCTTTTATTTTGAAGTAGTTATGTT 120

Db 663 WAWTTATTAATAAATTAATAATTAATAAAATTTTATWAAAAAATATTTTTTATATT 604

QY 121 CAATTCCTAATTAATGTATATTAATAAAATTCCTCGCAATATGTGAGAGGACCTCAT 180

Db 603 AATWAAATTTATATWATTAATAATWTTTAAATTTTAAATTTTAAATTTATTTT 544

QY 181 AATATTTGTCATATGGAATGAGCAGATAATAAAGATTATAGCTTTTCTTTGTCAAAGG 240

Db 543 TTTTWTATWAWWTWAAAWAATWAWWWAAAAATWWWWTTTTTTTWTATWATAWTT 484

QY 241 AGACTCAATATCTTTACTCTTTTCATGAGGACATTTGTG 277

Db 483 TTTTWWAATTTWKKKKKAWADTGAGARATKTDKK 447

RESULT 10

CNS000Z0C

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence T7 end of BAC:
BACN01C19 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.

Plasmid Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -

http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES

source

1..1101

/organism="Drosophila melanogaster"

/plasmid="pBelOBAC11"

/db_xref="taxon:7227"

/clone_lib="DrosBAC"

/clone="BACN01C19"

/note="end : T7"

BASE COUNT 354 a 173 c 149 g 278 t 147 others

ORIGIN

Query Match

Best Local Similarity

Matches 66; Conservative

69; Mismatches 87; Indels

0; Gaps

0;

Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999).

BASE COUNT 222 a 127 c 68 g 436 t

Query Match 6.0%; Score 51.8; DB 245; Length 853; Best Local Similarity 50.2%; Pred. No. 0.077; Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 444 TAAATCTAAACAGAGATTCGCATAGTTGGTGAATCTTTTAAACATCCCAATT 503
Db 425 TTAATATATTAACATAACACATTTTCAGAAATAAGAATGAATATTTAAACAATCAAAGT 366
QY 504 CATAGGCTTATAAATTAATCTGTATATTTTATCAACGAATCTGCCAGTTGCTTTCGTG 563
Db 365 GTTGATTTTAAATATTTTGTTCACATAAATTCATAAACAATATATTTTGATATATAT 306
QY 564 ATGCATAGAAAGATATAAAGAAAGAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
Db 305 ATAAATGGAATATTAAAGGATATACATACTCAAGAACTAATAAAGAAAGAAATAAAAAG 246
QY 624 GAAGCTCTGTTATAATGGGTCATGTAGATGGAAGAAGTATCTACATAAGCAGAGG 683
Db 245 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAAG 186
QY 684 AAGAGAAATGAAATA 698
Db 185 AAAGAATAATAGA 171

RESULT 15
LOCUS A2686179/c 868 bp DNA GSS 14-DEC-2000
DEFINITION ENTJ14LTR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION A2686179
VERSION A2686179.1 GI:11823325
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 868)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 74
High quality sequence stop: 779.
Location/Qualifiers
1. 868
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db.xref="taxon:5759"
/clone.lib="Entamoeba histolytica Sheared DNA"
/note="vector: pHOS1; Site 1; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a

FEATURES
source

tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999).

BASE COUNT 214 a 128 c 82 g 444 t

Query Match 6.0%; Score 51.8; DB 248; Length 868; Best Local Similarity 50.2%; Pred. No. 0.077; Matches 128; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 444 TAAAAATCTAAACAGAGATTCGCATAGTTGGTGAATCTTTTAAACATCCCAATT 503
Db 419 TTAATATATTAACATAACACATTTTCAGAAATAAGAATGAATATTTTAAACAATGAAAGT 360
QY 504 CATAGGCTTATAAATTAATCTGTATATTTTATCAACGAATCTGCCAGTTGCTTTCGTG 563
Db 359 GTTGATTTTAAATATTTTGTTCACACATAAATTCATAAACAATATATTTTGATAATAT 300
QY 564 ATGCATAGAAAGATATAAAGAAAGAAAGCTCAAGAACTCATAAAAACCCACACAATGT 623
Db 299 ATAAATGGAATATTAAAGGAAATATACAACTACTCAAGAACTAATAAAGAAAGAAATAAAAAG 240
QY 624 GAAGCTCTGTTATAAATGGTGGCTGCAATCTAAGATGGAAGAAGTATCTACATAAGCAGAGG 683
Db 239 AATGAATAAATGAAGATAAATTTGATAAATAATACAAGAAAGAAATATAAACGACAAAAG 180
QY 684 AAGAGAAATGAAATA 698
Db 179 AAAGAATAATAGA 165

Search completed: November 20, 2001, 15:31:57
Job time: 3966 sec